8-2015

Elucidation of Conformational Switching Mechanisms of Sensing Proteins by Molecular Dynamics Perturbation Studies

Quentin Ramon Johnson
University of Tennessee - Knoxville, qjohnso3@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Quentin Ramon Johnson entitled "Elucidation of Conformational Switching Mechanisms of Sensing Proteins by Molecular Dynamics Perturbation Studies." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Life Sciences.

Tongye Shen, Major Professor

We have read this dissertation and recommend its acceptance:

Jerome Baudry, Elias Fernandez, Hong Guo, Gregory Peterson

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Elucidation of Conformational Switching Mechanisms of Sensing Proteins by Molecular Dynamics Perturbation Studies

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Quentin Ramon Johnson

August 2015

Acknowledgments

Throughout my life I have been challenged again and again, scholastically, financially, physically, and mentally; through it all, I have depended on the support of my God, my friends and my family to keep my head high and my heart light. First, I give all the glory to God, for without him I would simply cease. Next, I thank my parents, George and Sylvia Johnson for bringing me into this world and providing me with the support, skills and knowledge so that I could be something great. To my sister, Monique, thank you for clearing the path first, so that my journey was easier, you mean more to me than you know. Then, there are my cousins, Courtney and Corbin Johnson, who were more like brothers to me growing up. While I have no biological brothers, you introduced me to the ideal of brotherhood, this virtue has served me well in forging lasting relationships throughout my life. In this life, friendship is essential to the soul, and no other entity has taught me this lesson better than Omega Psi Phi Fraternity Incorporated. I am extremely grateful to the my fraternity brothers that put me through the gauntlet, so that I could emerge

as a better man. There are too many names to explicitly thank everyone, but I am grateful to all my fraternity brothers. I especially thank the members of the Zeta Theta chapter for the guidance, tough love, and friendship they bestowed upon me. Finally, I must acknowledge those closest to my heart. To the light of my life, my lover and best friend, my one and only love in this life, my wife Esther Maria Johnson, thank you for all your love, support, and limitless compassion. Also, I thank my daughters Julissa Nicole and Khaleesi Maria Johnson for unknowingly providing me with motivation to achieve more in life. One day, you will understand how a child can make a parent better in every way.

Abstract

Sensing proteins are a subclass of switchable proteins, which are biologically designed with multiple, purposeful, low-free-energy states and can interconvert between these states in the wake of some environmental perturbation. Note, that this phenomenon is no small feat. This is a preprogrammed response for regulatory purposes that requires no cognitive action and is reversible as the environment returns to normal. Sensing proteins often switch between active and non-active states, closed and open conformations or other particular dichotomous states. Therefore, understanding the mechanism by which these proteins sense a specific perturbation and how they switch between conformations is paramount. Addressing these issues can lead to the ability to control protein efficiency; by either direct manipulation of the protein itself to increase its sensitivity to the environment or by quantifying the optimum environmental condition for the desired function. Here molecular dynamics simulations and multi-scale analysis tools are used to elucidate the switching mechanisms of sensing proteins.

Table of Contents

List of Figures

Abbreviations

PAL...*Pseudomonas aeruginosa* lipase

PCA...principal component analysis

RMSD...root-mean-square deviation

2D-RMSD...two-dimensional root-mean-square-deviation

MD...Molecular Dynamics

fs...Femptosecond

ns...Nanosecond

CAMERRA...Computation of Allosteric Mechanism by Evaluating Residue-residue Associations

PMF...Potential of mean force

IgA...Immunoglobulin A

GalNAc...N-acetylgalcatosamine

Gal...Galactose

NeuAc...Sialic acid

RXR...retinoid X receptor

TR...thyroid hormone receptor

9C...9-cis retinoic acid

T3...3,3',5 triiodo-L-thyronine

CAR...constitutive androstane receptor

NR...nuclear receptor

PCA...principal component analysis

PC...principal component

DOF...degree of freedom

Chapter 1

Introduction

Proteins are commonly referred to as the workhorses of nature, since they perform many of the tasks that are essential for life. They are integral in transcription, translation, metabolism, respiration, and countless other biochemical processes [1, 2]. Indeed, proteins are an irreplaceable cog in the machine of life, and their inherent design is exceptional since they are biologically programmed to automatically carry out a particular function [3]. This inherent design, or primary structure, carries the blueprint to properly fold the protein into a stable and functional conformation (given suitable environmental conditions) [1, 4, 5].

Researchers have shown that protein structure is directly linked to function, and it is well known that under favorable conditions proteins usually adopt a certain low-free-energy state, which is termed the native state [6, 7]. However, a special

class of proteins is not restricted to a single free energy basin, but actually exhibit multiple basins [8]. These "switchable" proteins are able to switch their preferred conformation as they move from one basin to another on their energy landscape. Often, the cause of this conformational switch can be traced back to specific environmental perturbation that causes a cascade of alterations in intra-protein and inter-protein interactions that inevitably change the global conformation [4, 9]. An example of this is the bacterial lipase from *Pseudomonas aeruginosa* which adopts a closed state (inactive) in an aqueous environment, but switches its preferred conformation to an open one (active) upon entering a hydrophobic environment [10]. Here, the solvent that surrounds the protein has a marked influence on its energy landscape and thus its most probable structure.

Sensing proteins are a subclass of switchable proteins, which are biologically designed with multiple, purposeful, low-free-energy states and can interconvert between these states in the wake of some environmental perturbation. Note, that this phenomenon is no small feat. This is a preprogrammed response for regulatory purposes that requires no cognitive action and is reversible as the environment returns to normal [11]. Sensing proteins often switch between active and non-active states, closed and open conformations or other particular dichotomous states. Therefore, understanding the mechanism by which these proteins sense a specific perturbation and how they switch between conformations is paramount. Addressing these issues

can lead to the ability to control protein efficiency; by either direct manipulation of the protein itself to increase its sensitivity to the environment or by quantifying the optimum environmental condition for the desired function.

There are several types of sensing proteins, those that are sensitive to temperature, pressure, solvent, pH, small molecules, etc. However, it is not likely that these proteins operate via a universal switching mechanism [12]. This raises the question, can different switching mechanisms be commonly defined or is each unique? In addition, if there is no universal mechanism then can a set of sub-mechanisms be defined so that all temperature sensing proteins operate the same, or all pressure sensors, and so on? It is even conceivable that the secret lies at the protein level. Do all lipases follow a uniform switching mechanism or all membrane proteins, etc? These questions can only be answered by delving into the problem at multiple levels and by comprehensive review of the field.

Sensing proteins offer a unique scientific challenge, as observing a switch between conformations is not a simple task. Environmental perturbation can bring about millions to billions of subtle changes in intra-protein, inter-protein and protein-solvent interactions that will ultimately lead to a new stable conformation [13]. To further complicate matters, these interactions can be formed or broken on a picosecond (ps) time-scale [14], therefore a comprehensive understanding of these switching mechanisms requires a sharp temporal and spatial resolution. To address

this, molecular dynamics (MD) simulation can be employed as a complimentary method with experiments in order to satisfy this need.

MD simulation is an *in silico* procedure that utilizes Newtonian mechanics to predict the movement of molecules with atomistic detail and femtosecond (fs) resolution [15]. Almost half a century ago the first biological atomistic MD simulation was performed, this pioneering exploit focused on the BPTI protein and employed the same basic procedure still used today [5]. Of course, the field has vastly matured since then, current force fields are more accurate and more extensive to include exotic molecules such as saccharides, lipids, nucleic acids and unique ligands [16]. Also, the hardware has exponentially improved to increase simulation run time and speed. For example, early simulations were performed on a picosecond time-scale, while today scientists are pushing the microsecond boundary and hope that millisecond simulations will soon be a routine practice. Simulations are used for a variety of purposes, such as, to provide conformational information, information on dynamics, to compute free energy, or binding energy.

Given the atomic nature of MD simulation, it can be very useful in providing a detailed picture of the inner workings of switching mechanisms. Additionally, the scrutinous temporal resolution allows one to track pertinent information on a femptosecond timescale. These characteristics provide an edge to simulation over many other techniques in examining switching mechanisms. However, the importance

of experimental compliment should not be overlooked as these studies can provide starting structures for simulation, evidence of interesting conformational regions, and validation for simulation results [17].

Indeed, MD simulation is quite adept at addressing switching mechanisms, but due to the basics of molecular dynamics simulation, these types of studies can be difficult to perform directly. Consider this: a simulation is setup with a concrete set of parameters and conditions; it is often impossible to reset these conditions mid-stride without disrupting the integrity of the system. For example, a sharp increase in simulation temperature will cause the atomic forces to vastly increase, this will result in large changes in atomic velocities and positions, which will ultimately lead to a cascade of unphysical forces in the system and "crash" the simulation, (similar issues exist for other environmental factors). Therefore, direct manipulation of the environment during the course of a simulation can be troublesome, so a clever scheme must be devised to indirectly model these situations. In this work, I detail several of these molecular dynamics simulation methods and data reduction analysis methods for the elucidation of different switching mechanisms. I show that this type of phenomenon can be accurately addressed using MD simulation and that the results gleaned from these theoretical studies are practically relevant.

Chapter 2

Theory of Molecular Dynamics Simulation

In this age of supercomputing, an array of methods to perform molecular dynamics (MD) simulations exist, such as traditional all-atom simulation which provides a detailed spatial resolution, coarse grained simulation strategies which allow larger systems to be studied using smaller amounts of computational resources, and accelerated MD which aids in the observation of long time-scale processes [18]. However, all these diverging strategies stem from a common theory that given the correct approximations and parametrization the underlying forces that govern molecular interactions can be accurately estimated. Here, the core of that theory is expounded upon.

The strategy behind molecular dynamics simulation is to take a static depiction of a system x and utilize Newton's equations of motion to estimate the position of the system at a given time t [18]. Incumbent in this calculation is the knowledge of how each particle in the system will interact with every other particle, this is known as the force field. The combination of these three things produces a predictive model of how atoms change position over time which is recorded as a series of coordinates, termed a simulation trajectory. From this trajectory many questions about the physical behavior of the system can be addressed.

2.1 Force Field

Molecular dynamics simulation relies on the principle that molecular motions are controlled by a few distinct forces, a comprehensive description of these forces is referred to as the force field. From a base quantum mechanical view, all forces are rooted in the electrostatic and spin coupled interactions between electrons and nuclei, therefore quantum mechanical calculations are routinely used to derive parameters for the set of forces used in MD simulation. These parameters are then applied via the force field to all atoms and thus the total potential energy can be calculated. A typical force field is displayed in the equation below [19].

$$E(\mathbf{r}^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + cos(n\omega_i - \gamma_i)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \tag{2.1}$$

Here, the total energy potential is made up of four terms, Σ_{bonds}, Σ_{angles}, and $\Sigma_{dihedrals}$ represent the covalently bonded interactions, and the final term is for the non-bonded interactions which are made up of the van der Waals short ranged repulsion and attraction with a Lennard-Jones potential and long-ranged coulombic (electrostatic) interactions. This common force field is more than adequate for routine simulation work, but is not intended to address situations where chemical bonds are broken or formed. Modified versions of the basic force field have been developed to accomplish this task, such as QM/MM simulation techniques [20]. Additionally, this basic force field does not account for dynamic atomic polarizability, which is arguably important for the simulation of highly polarizable molecules such as carbohydrates [21]. There are also special force fields available to satisfy this feature [22]. However additions to the base force field do come at a price, as the inclusion of these special parameters increases the computational cost by a large degree. Therefore, the vast majority of MD simulation studies are done with the common force field or some small variant of it, especially when conformational sampling is of issue.

2.2 Equations of Motion

As mentioned above, the force field is used to compute the potential energy, which is a function of the coordinates of the atoms in that system. Atomic nuceli basically behave as classical particles and their dynamics can be predicted by solving Newton's equations of motion.

$$\mathbf{F_i} = \frac{-\partial U}{\partial \mathbf{r_i}} = m_i \frac{d^2 \mathbf{r_i}}{dt^2} \tag{2.2}$$

$$\mathbf{a_i} = \frac{d^2 \mathbf{r_i}}{dt^2} = \frac{d \mathbf{v_i}}{dt} \tag{2.3}$$

Here U is the potential energy of the system at a given time t, F_i is the force on an atom i, m_i, r_i, a_i and v_i are the mass position, acceleration and velocity of i, respectively. Ideally, this calculation would be done over an infinitely small time step in order to accurately characterize the continuous motion of each atom. In reality, a finite-difference method is utilized to numerically integrate Newtons equation of motion to compute atomic coordinates and velocities. One such method that is commonly employed in MD simulation is the Verlet algorithm [23]. The Verlet algorithm relies on a basic Taylor expansion to update the position r at a new time δt. The final form is:

$$\mathbf{r_i}(t + \Delta t) = 2\mathbf{r_i}(t) - \mathbf{r}_i(t - \Delta t) + \mathbf{a_i}(t)\Delta t^2 + O(\Delta t^4) \tag{2.4}$$

The velocities can be computed from the coordinates by using the following formula:

$$\mathbf{v_i}(t) = \frac{\mathbf{r_i}(t + \Delta t) - \mathbf{r_i}(t - \Delta t)}{2\Delta t} + O(\Delta t^2) \tag{2.5}$$

Due to the apparent simplicity of this method it is perhaps the most widely used for calculations of this nature, although many improvements have been introduced over the years to reduce error. It is also important to note that MD simulation requires a suitable choice for the integration step to save computational time without losing accuracy. As Newton's equation of motion is integrated, the highest frequency that occurs in the simulation system is the limiting factor of the chosen time step.

2.3 Constraint Algorithms

Hydrogen, being the lightest known atom, is usually involved in bonds that vibrate with the highest frequency, therefore they determine the maximum usable time step. To increase this integration time step constraint algorithms have been introduced to restrict certain degrees of freedom (such as bond lengths and angles). Use of these algorithms allow the time step to be increased to a typical value of 2 fs. A well accepted algorithm for this task is SHAKE [24].

2.4 Periodic Boundary Conditions

Biological systems often contain billions of molecules, which can freely interact with each other. This near-infinite size system is impossible to simulate using modern computers (even supercomputers). To address this issue, MD simulations can model the behavior of these molecules by using a subpopulation and confining this subpopulation to a small three-dimensional box, called a unit cell. To enhance the realism of this unit cell in comparison to real life, periodic boundary conditions (PBC) are often employed. This means that one unit cell is surrounded by periodic images of itself, and when an object passes through one side of the unit cell it instantaneously enters through the opposite side with an equal velocity. In this way the number of molecules in the system remains constant. While the obvious choice for the shape of this box is a cube, this can lead to the addition of solvent molecules in the corners, which are quite far from the central biomolecule of interest. The addition of these unnecessary solvent molecules greatly increases the computational cost. Therefore, a common, more efficient alternative is a truncated octahedron.

2.5 Non-bonded Interactions

Periodic conditions, while effective in reducing computational cost by decreasing the amount of molecules needed for realistic simulation, can be troublesome for

calculating non-bonded interactions since interactions between particles in different boxes are allowed. Consequently, the amount of pair-wise interactions increases drastically. This is due to the fact that in addition to determining the interactions between molecules in the "real" unit-cell, interactions between the real and image cells must also be considered. Therefore, the bulk of computational energy is expended calculating the non-bonded van der Waals and electrostatic potentials. Usually, the van der Waals interaction is described by the Lennard-Jones potential and the electrostatic is described by interaction between point charges as follows Coulomb's law. To combat the high computational cost of calculating non-bonded interactions a cut-off radius of roughly 1 angstrom is routinely employed for short range interactions. It should be noted that the use of this cut-off requires long range correction for the dispersion term of the Lennard-Jones potential, although no correction is needed for the repulsion term as it decays quickly with respect to r.

2.6 Temperature Coupling and Pressure Coupling Methods

In MD simulation there is often a need to regulate the temperature and pressure variables rather than energy and volume. To produce this effect the equations of motion need to be modified with temperature coupling and pressure coupling meth-

ods. The need to study properties at isobaric and isothermal conditions has led to the development of several methods to keep these parameters constant. Note that the temperature T of a system is related to the kinetic energy K by the following equation.

$$K = \sum_{i=1}^{N} \mathbf{m_i} |\mathbf{v_i}|^2 / 2 = N_{df} \mathbf{k_B} T / 2 \tag{2.6}$$

The pressure p of a system is related to the box volume V_{box} and the kinetic energy of the system by the following equation.

$$P = \frac{2}{3V_{box}} \left[K + \frac{1}{2} \sum_{i<j} \mathbf{r_{ij}} \cdot \mathbf{F_{ij}} \right] \tag{2.7}$$

A common method for temperature and pressure regulation is the Berendsen method. [25] Berendsen temperature coupling the system is weakly coupled to a heat bath with some temperature. The temperature of the system is corrected such that the deviation exponentially decays with some time constant τ. For large systems, this approximation yields roughly correct results for most calculated properties. Here the change in temperature is proportional to the temperature difference between the heat bath and the system with a coupling strength determined by the coupling parameter τ. This is defined by the following equation.

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \tag{2.8}$$

Berendsen pressure coupling is coupled to an external pressure bath with reference pressure P_0. The coordinates and box vectors are rescaled every step. The system pressure os regulated with first-order kinetics. Berendsen pressure coupling is achieved using the following equation.

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_p} \tag{2.9}$$

Other common coupling methods are Langevin, Nosé Hoover, and Parrinello-Rahman [26,27]. When deciding which method to use several factors are important, such as, system size, efficiency, and accuracy.

Chapter 3

Solvent-Dependent Gating Motions of an Extremophilic Lipase from *Pseudomonas aeruginosa*

Solvent-dependent gating motions of an extremophilic lipase from *Pseudomonas aeruginosa*

Quentin R. Johnson,[1,2] Ricky B. Nellas,[2,3] and Tongye Shen[2,3]*

[1]UT-ORNL Graduate School of Genome Science and Technology, Knoxville, TN, 37996

[2]Center for Molecular Biophysics, Oak Ridge National Lab, Oak Ridge, TN, 37830

[3]Department of Biochemistry and Cellular & Molecular Biology,

University of Tennessee, Knoxville, TN, 37996

Abstract

Understanding how organic solvent-stable proteins can function in anhydrous and often complex solutions is essential for the study of the interaction of protein and molecular immiscible interface and the design of efficient industrial enzymes in nonaqueous solvents. Using an extremophilic lipase from *Pseudomonas aeruginosa* as an example, we investigated the conformational dynamics of an organic solvent-tolerant enzyme in complex solvent milieux. Four 100-nanosecond molecular dynamics simulations of the lipase were performed in solvent systems: water, hexane, and two mixtures of hexane and water, 5% and 95% (*w/w*) hexane. Our results show a solvent-dependent structural change of the protein, especially in the

region that regulates the admission of the substrate. We observed that the lipase is much less flexible in hexane than in aqueous solution or at the immiscible interface. Quantified by the size of the accessible channel, the lipase in water has a closed-gate conformation and no access to the active site, while in the hexane-containing systems, the lipase is at various degrees of open-gate state, with the immiscible interface setup being in the widely open conformation ensembles. The composition of explicit solvents in the access channel showed significant influence on the conformational dynamics of the protein. Interestingly, the slowest step (bottleneck) of the hexane-induced conformational switch seems to be correlated with the slow dehydration dynamics of the channel.

The structural stability of proteins involves a delicate balance between the favorable internal interaction among residues of the protein and the protein-solvent interaction. Many proteins optimize their structural stability in an aqueous environment. [1] When submerged in organic solvents or placed at immiscible interfaces, their native structures are likely to be weakened and their activities diminished. But for a few families of proteins that interact directly with an organic substrate (at the interface), the desired reaction environment may be more complicated. For some of these biomolecules, their functions are kept in a seemingly anhydrous environment due to the presence of a small amount of hydration water. [28] However, there are exceptional natural proteins, from surfactins [29] to lipases [30, 31], that operate efficiently in a water-oil interface or in organic solvents, especially some enzymes of microbial origin that thrive in deep-sea mud. [32] These proteins are structurally stable and biosynthetically active at an organic solvent interface.

In this work, we focus on one such important extremophilic enzyme, a lipase from *Pseudomonas aeruginosa*, which is inactive in an aqueous milieu and activated at the water-oil immiscible interface. [10, 30, 33] Lipases are proteins that assist in the forming and breaking of ester bonds of glycerol esters. [34, 35] They are known to play important roles in biochemical pathways, such as transport processes and stereo-selective biotransformations. [35] Besides these natural biological functions, lipases have a diverse set of industrial applications, from the break down

of unwanted oily components of cosmetics, detergents, and waste to the synthesis of esters. [36]

Lipases belong to the superfamily of α/β-hydrolase, [34] a group of enzymes (including proteases, esterases, lipases, and peroxidases) sharing the same α/β fold. [37] At the active site, these enzymes share the catalytic triad motif for breaking covalent bonds. Apparently, this well-studied enzymatic mechanism is quite efficient once the substrate reaches the active site. [34] The discerning factor among hydrolases is substrate selectivity, which begins at the precatalytic step. The concealed catalytic cavity and the switchable gate enable conformational control of ligand admission. [37,38] Indeed, lipases obtained from different species, from fungus (*Candida antarctica*) to bacterium (*Burkholderia cepacia*), share this canonical fold and the catalytic triad. [39,40] As shown in Fig. 3.1a, the lipase from *P. aeruginosa* folds with its active site buried at the center. There is a 15 Å long gated channel that leads to the ovoid shaped active site cleft. [41] The main feature of this hydrophobic channel is the 'lid' region that regulates entry of substrates and solvent molecules to the gorge. Movement in this region is critical for the enzyme activity. [41]

A plethora of crystal structures of different lipases at various binding states have been reported in an effort to understand how this channel operates. [40–46] They reveal that this channel exists in different conformational states. The ligand-bound form is often associated with the channel being relatively wide open (open state),

while the apo form of the lipase presents a closed channel without ligand access (closed state). Several aspects of this conformational gating mechanism are still missing. Firstly, it is important to provide an ensemble view of probable states of the conformation of the access channel throughout the gating motion, not just the qualitative description of dichotomous states (open vs. closed). Secondly, the mechanism of opening and closing the channel is solvent-sensitive, and it is important to study the conformational dynamics at realistic complex solvent interfaces.

Computational studies on the aforementioned lipases in solution have also been carried out to obtain insight on how the solvent facilitates the conformational shifts of lipases. [39, 47–55] Several molecular dynamics simulations have discovered important motions of lipases, especially the movement of the gate triggered by the hydrophobicity of the environment. [47] To our knowledge, all previous simulations of lipases are about 35 ns or much less. As shown in a previous study [39] and in the results presented in this work, it may take lipases as long as 30 ns to repel water molecules inside the gorge and to make the conformational switch to close the gate. Thus, longer time scale simulations are required to get better statistics. A more quantitative analysis of the gated access channel (as oppose to the open/closed binary description of the end points) will facilitate a comparison of the conformation ensemble in a more precise manner, as we will illustrate below using the gorge radius analysis.

Several studies on how the open-close conformational switch of lipases may work in various solvent systems have been reported. For *Burkholderia cepacia* lipase (BcL), when in aqueous solution, an initially open conformation gradually closes. Conversely, when in toluene (also in octane or water-octane (50/50) interface [50]), an initially closed BcL opens gradually. [39] For *Bacillus sp.* (Lip42), in pure water and in 100% dimethyl sulfoxide (DMSO), the access channel seems to be unperturbed while in 60% (*v/v*) DMSO, distortion of conformations (accompanied by formation of nonnative H-bonds) is observed. [51] For *P. aeruginosa* (PAO1), when in water, an initially open structure becomes closed, and starts to open again when moved to a water-octane (50/50) interface. [47] Note that all of these previous studies are performed either in a pure solvent or a relative "bulky" immiscible interface, such as the water-octane (50/50) interface. The effect of the presence of a small amount of water or organic moiety in the solvent mixture on the conformational transitions of the lipase is much less examined. Most practical interfaces encountered by lipases may not be an ideal case. Rather, we are interested in less well-formed molecular interfaces with a small number of organic solvent molecules surrounded by water or vice versa. Such setups (with less surface area between the immiscible interface) carry less perturbation due to surface tension energy and thus minimizes the effects of bulky interface.

Lipase from *Pseudomonas aeruginosa* (PAL) strain LST03, has already been

characterized in terms of sequence and activity. [32, 56, 57] It has been reported to show unusual stability and activity in organic solvent systems. The succeeding discussions will focus on the statistical analysis of the gating motion of the lipase in four water-hexane solvent compositions and the role of gorge solvent molecules on the conformational dynamics of the protein.

Systems and Methods

Using the lipase from *P. aeruginosa*, we performed a series of molecular dynamics simulations in four explicit water-hexane solvent environments: 100% hexane, 95% hexane, 5% hexane, and 0% hexane (pure water). For convenience, these four systems are labeled as "100H", "95H", "5H", and "0H", respectively. Although the sequence of PAL-LST03 is known, its crystal structure is not yet available. Thus, we turn to its relative, the lipase from *P. aeruginosa* (PAO1). This lipase has 285 residues and its crystal structure has been reported in protein data bank as PDB ID: 1EX9. [41] It only differs from PAL-LST03 by a single residue (a very subtle change of 1EX9:Val^{130} to Ile in helix α5). A separate study demonstrated that this substitution does not affect the conformations of the peptide (α5 region) in different solvent conditions (data not shown).

The crystal structure, PDB:1EX9 [41] (with ligand removed) was used as the

starting structure for all simulations. Systems were prepared using the xleap module in AMBER 10. [19] AMBER 99SB force field was used for the protein while TIP3P [58] was used for water. The organic solvent hexane was modeled using the antechamber module of AMBER 10. [19] For the construction of the 0H system, the protein is immersed in a periodic box containing 10 465 water molecules. For the 100H system, the protein is solvated in 2 175 hexane molecules. The initial state of the 5% (4.86%) hexane mixture (5H) system was constructed by first solvating the protein with 113 hexane molecules and then adding 10 591 water molecules on the outer layer. The initial state of the 95% (95.23%) hexane mixture (95H) system was constructed by first solvating the protein with 200 water molecules then adding 836 hexane molecules on the outer layer. Finally, three sodium ions were added to neutralize each system.

The presence of a tightly bound calcium ion is critical to the structural integrity of the lipase. The coordinated metal ion (Ca^{2+}/Zn^{2+}) enhances the thermostability of the protein and its absence causes major structural changes. [44,52,53] For PAO1, Ca^{2+} stabilizes the loop containing His^{251}, a part of the catalytic triad. [41] Since the metallic bonding aspect of the calcium is difficult to incorporate, we simply fix the positions of calcium and the four neighboring oxygens (Asp^{209}:O, Asp^{253}:O, Gln^{257}:O, Leu^{261}:O) using the protocol described in NAMD2.7. [59]

For each system, minimization, heating, and equilibration (for approximately

10 ns) was conducted and followed by a 100 ns molecular dynamics production run using NAMD2.7 package. [59] Simulation pressure and temperature are kept constant at 1.01325 bar and 300 K, respectively. Langevin thermostat is used to regulate temperature. Long-range interactions are treated using the particle mesh Ewald method. All bonds involving hydrogen atoms are constrained using the SHAKE algorithm. A time step of 2 fs is used to integrate the equations of motion. Frames are collected at 1 ps intervals. Data collected from the 10 ns equilibration run was disregarded and analysis was performed on the final 100 ns molecular dynamics production data.

The program HOLE [60] was used here to probe the gorge radius of the lipase. This software has been routinely employed to measure the size of ion channels and cavities of protein structures. [60–62] We performed the analysis on each snapshot and obtained a global measurement of the openness of the access channel that leads to the active site. This procedure does not measure a particular distance between arbitrarily defined parts of the channel. Rather, it gives the size of the bottleneck of the channel, which is dependent on the snapshots. This method requires an input of atomic coordinates, the point of origin of the gorge (cpoint), and a defined gorge direction (cvect). [60] A library of atomic radii is included in the HOLE program and was used to construct the geometry of the protein surface. For each conformation of the lipase, HOLE is able to calculate the effective size of the channel as a function

of the semiautomatic found path, $\rho(\vec{r})$. The path is a set of coordinates linked from the active site to the entrance. We then further defined the minimum radius along this path as the bottleneck.

The two crucial factors for this probe are the cpoint and cvect. We first calculated the center between four residues, (Gly^{24}, Phe^{214}, Leu^{138}, and Phe^{117}), that surround the opening of the gorge in all cardinal directions. We then find the average position (cpoint) between this center point and the position of Ser^{82}. Similarly, the cvect originates at the cpoint and extends along the z axis towards the surface of the protein.

Results and Discussion

As mentioned previously, the initial structure of the lipase for all simulations was that of the crystal structure complexed with a ligand (removed for this study) at the active site PDB:1EX9. [41] This structure was deemed to represent an open and active conformation. Lipase are widely believed to be activated only at an oil-water interface. [39] Thus, we are interested in whether the lipase in each solvent condition will remain in the initial conformation or explore others, such as the closed conformation that features a closed channel with no ligand access. Practically, this closed conformation is associated with a constricted gate via the displacement of

the $\alpha 5$ helix towards the $\alpha 8$ helix. [47] Thus, one particular interesting region is the $\alpha 5$ helix (residues 126-147) because the movement of this helix seems to govern the opening of the gate. We also pay attention to $\alpha 8$ (residues 210-222) because it has previously been proposed to affect the gating mechanism as the second lid. [47] Another important aspect of this work is the solvent analysis as we will relate the dynamic motion of the protein to that of the solvent molecules.

3.1 PCA Analysis

Before going into the local motions, we examine the large scale motions of the protein by the principal component analysis (PCA) of the covariance matrix of the C_α coordinates. Measured by eigenvalues of PCA, PC 1 and PC 2 (the most dominant PCs) contribute 54.1% and 15.3% respectively to the overall motion (normalized by the trace of the covariance matrix). Together, the ten most dominant PCs contribute 95% to the overall motion. In Fig. 3.1b, PC 1 and PC 2 represent the top two principal components calculated from a mixture of 140 000 snapshots of the 0H and 100H trajectories (70 000 each). Then the final 70 000 snapshots of each system were projected onto PC 1 and PC 2. The color legend (each color contains 10 000 snapshots, representing 10 ns simulation) shows the time evolution of the two principal components. From Fig. 3.1b, we observed that PC 1 clearly demonstrates that

the lipase has a different conformational ensemble when in an aqueous solution and when it is in hexane-containing solvent systems. Here, 0H exhibits a negative PC 1 while all the hexane-containing systems have a positive PC 1. Also, 0H shows more than double the amount of variation along PC 2 compared to the hexane-containing systems. Whether the system is settled down in a free energy landscape basin or not was roughly monitored by the motion of the protein in ten-nanosecond blocks, i.e., how conformational ensemble of each time block is different from that of the previous block. Judged by the time evolution of global motions projected onto PCs 1 and 2, we observed that the 0H and 95H systems take roughly 50-70 ns to begin to converge while the 5H and 100H systems take about 20 ns or less to show convergence.

Upon further structural examination (after convergence), the lipase from 0H appears to have a closed channel, while the lipase in the hexane-containing systems have an open channel. For the current setup, it takes tens of nanoseconds for a lipase from crystal structure conformation to settle down to the closed conformation ensemble when it is immersed in aqueous solution. Interestingly, the systems that take the longest to converge were constructed with water molecules initially placed on the surface of the protein, whereas systems that converge in less than 20 ns were constructed with hexane molecules in the first solvation layer. In addition, from the areas of the final 10 ns color block (gray) in the PCA plot, the flexibility of the

lipase in increasing order is 100H, 0H, 95H, and 5H. This indicates that basins of free energy landscape of water-containing systems are flatter which facilitate conformational exploration of the lipase while pure hexane confines this exploration. It would also be interesting to quantify this organic solvent contribution to the energetic roughness [63] of the protein landscape in complex solvent systems in the future.

3.2 Flexibility of Lipase Influenced by Solvent Environments

Figure 3.1c shows a plot of the two dimensional overall root mean squared deviation (2D-RMSD) of each system. Here the x and y axes represent the mean structures of the 0H and 100H systems, respectively. The 2D-RMSD is calculated from the trajectory of each system and projected onto the average structures of the lipase in 0H and 100H. As expected 0H and 100H systems show a low RMSD value on their respective axes. However, both 5H and 95H show an RMSD greater than 3 for each axis. It is apparent that while each initial lipase structure is identical, the unique protein-solvent interactions in each simulation lead the lipase along a pathway towards a specific local low energy conformation. Remarkably, lipases solvated in the water-oil mixtures seem to converge to similar structural ensembles

regardless of the difference of the overall composition of the mixture.

Since we have characterized the global motions of the lipase, we now zoom in on the local aspects of the access channel. First, we examine the local flexibility of the lipase in different solvent environments using the Debye-Waller factor (B-factor) which reveals the locality of an atom and can be calculated by using the mean-square fluctuation, B = $8\pi^2(\Delta r)^2/3$. Figure 3.1d shows the deviation from the mean at the residue level, calculated from the last 70 ns. The α5 helix (residues 126-147) consistently shows one of the highest B-factor values for all four systems, especially for 0H system. Also, the α5 shows a comparable B-factor in the two mixture systems, 5H and 95H. The systems (0H and 95H) that were constructed with water molecules in the first solvation layer show a similar B-factor for the α8, and the same holds true for the 5H and 100H systems with hexane molecules in the first solvation layer. This increase in fluctuation around the average structure for the α5 in the 0H system is contributed by the transient motion of gate closing, not the equilibrium fluctuation. Figure 3.1e shows the B-factor for each system calculated from the last 10 ns of the simulation where this transient motion is eliminated. Here we see that the 5H system is most flexible, 95H follows, then 0H, and 100H. This is consistent with the trend observed from the PCA analysis. Again, from PCA, overall RMSD and B-factor analysis, we observed that the rigidity of this protein is increased when it is immersed in hexane. Our observation of lower flexibility

in hexane media (than in water) is also consistent with study performed on another hydrolase system, crystal structure of chymotrypsin with hexane. [64] Many factors, from specific protein-solvent interactions to generic hydrophobic solvent effect, may contribute to the origin of this immobilization effect by hexane media. For example, a hexane environment does not provide solvent hydrogen bonds that would otherwise "lubricate" the protein motion.

3.3 Gating Motion of the Access Channel

By monitoring the gorge radius, we quantify the solvent-induced gating motions of the lipase. From the crystal structure, this gorge has an ovoid shape of radius 3 Å and dimensions 10 Å $\times$ 25 Å. [41] In the four solvated systems with ligand removed, the size and shape of this gorge changes to accommodate solvent molecules.

In Fig. 3.2a, the gorge radius is shown as a function of time, from which we observed that the gating motion of the lipase is highly active at the interface (5H and 95H systems). In the 100H system, the gorge radius of the lipase persistently fluctuates around 1 to 2.5 (reaches a maximum of 3 Å), which implies that there is very little structural perturbation on the initial configuration. In contrast, the gorge radius in 0H system slowly but steadily decreases from 3 to 1 Å. Using simple distance criterion, similar conclusions on gate closing in aqueous simulation of lipases

have been reported previously. [39,47] The radius of the gorge in 5H reaches nearly double that of the crystal structure (from 3 to roughly 6 Å), and 95H reaches a gorge radius of about 4 Å. This opening of the gorge is possibly the result of the complimentary forces that act on the protein in a complex interfacial environment. We will discuss the opening mechanism in detail below from the movement of the amphiphilic α5 perspective.

The distribution of gorge radius $p(r)$ shown in Fig. 3.2b reveals interesting bi-modal distributions, especially for that of the 100H system. The lipase has the highest probability of being in an open-state in 5H (broadest distribution) while least in 0H (narrowest distribution). The 5H and 95H having similar distribution at larger gorge radius denotes that the lipase favors an open conformation at an interface. Based on the distributions, it appears that an open-gate conformation is most likely achieved when small amount of organic moiety is present in an aqueous environment.

3.4 Solvent Molecules inside the Access Channel

In Fig. 3.3a, we monitor the number of solvent molecules found in the access channel as a function of time. This analysis is done by searching for solvent molecules within a radius of 10 Å from Ser^{82}. Note that in the crystal structure the distance

from the active site to the mouth of the access channel is about 15 Å. Several other cutoffs greater than 10 Å were tested but yielded significant inclusion of bulk solvent molecules (data not shown). The top three panels in Fig. 3.3a represent the number of hexane molecules found in the gorge throughout the course of the simulation. Notice that each hexane-containing system does allow roughly 5-10 hexane molecules into the gorge, but the 5H system clearly admits the greatest amount of hexane molecules into the gorge, as lipase of 5H exhibits the widest gorge opening (Fig. 3.2). Figure 3.3b displays the statistics of solvent molecules. Here, the number of hexane molecules in the 5H system shows the widest distribution. This indicates that a significant amount of hexane molecules are moving in and out of the gorge. The relatively constant amount of gorge hexanes in 100H may be correlated to the insignificant change in gorge size (Fig. 3.2). As shown in Fig. 3.3b, 100H has the narrowest distribution of gorge hexane molecules. The 95H system shows an initial increase in the amount of gorge hexanes, but it is less than the amount of hexanes in 5H, as the protein in 95H does not open as wide as that in 5H.

In Fig. 3.3a, the bottom three panels represent the number of water molecules found in the gorge as time evolves. Here, the 0H system initially has the greatest amount of water inside the gorge due to an initial void in the access channel left from the ligand removal. However, the water molecules are squeezed out as the gorge narrows due to its hydrophobic nature. During this process, the lipase in 0H

is slowly transformed from an initial open to a closed conformation. The widest distribution of gorge waters in 0H shown in Fig. 3.3b also indicates a substantial amount of water molecules being excluded from the gorge. In agreement with the PCA data (Fig. 3.1), the amount of water becomes relatively stable around the 70 ns mark as the protein relaxes into the close conformation ensemble. Both 5H and 95H show a modest increase in the amount of gorge waters. The 5H system shows more water molecules in the gorge as it has a more open structure. Figure 3.3 clearly shows that 5H accommodates the highest amount of solvent molecules. Thus, the molecular interface constructed by the gorge solvent molecules can be seen as being vital in the gating mechanism of the access channel.

3.5 The Role of the Lid Region in the Mechanism of Interfacial Activation

As previously stated, the lipase regulates gating via the motion of the lid region. This region is defined mainly by four helices: α4 (residues 113-120), α5 (residues 126-147), α6 (residues 155-162), and α8 (residues 210-222). Previous studies have shown that the channel can be closed either by movement of the α4 and α6 helices toward each other or by shifting α5 and α8 closer together. [47] We have found by local RMSD analysis (not shown here) that by a significant magnitude the mo-

tions of the other three helices (α4, α6, α8) are negligible compared to that of amphiphilic α5. So, we focus on the relative motion of α5 and α8 and how this constriction is related to the opening and closing of the channel. The distance between α5 (Cα of Ile^{159}) and α8 (Cα of His^{215}) reported in Fig. 3.4a shows that in the 0H system α5 constricts towards α8 as the simulation proceeds, from roughly 23 Å to 17 Å. Also, it shows that the inter-helix distance actually reaches more than 20 Å in all the hexane-containing systems. Here 5H yields both the highest average value and highest maximum value.

The constriction of the lid region in 0H may signify that the closure is driven by the hydrophobicity of the inner "core" face of the amphiphilic α5 (Fig. 3.4b). The interaction of the aqueous solvent and α5 may push the helix inward. [65] Conversely, the highly hydrophobic solvent environment of 100H interacts with α5 differently, and α5 shows no such movement towards α8. Although 5H and 95H are both good candidates for interfacial activation, the 5H system (Fig. 3.3) allows more solvent molecules into the gorge and displaces α5 away from α8. In contrast to the pure systems, the mixtures present the environment for ideal interfacial activation wherein complementary forces may act on the amphiphilic helix.

Conclusions

Four sets of 100 ns simulations of the lipase were performed in aqueous solution, hydrophobic solvent (hexane), and two water-oil mixtures (5H and 95H) with molecular immiscible interfaces. Due to the long time-scale and unique solvent composition used in these simulations we are able to quantitatively observe the gating motions of this lipase in detail. We found that the relaxation time is quite long (tens of nanoseconds) for two of the systems that have water placed inside the channel as the initial states (0H and 95H systems). The bottleneck of the gating dynamics seems to be the repelling of water by either constricting the gorge or replacing water with hexane. As supported by both PCA and 2D-RMSD, the lipase assumes distinct conformational states when exposed to different solvent systems studied here. The simulation in pure hexane produces the least structural deviation from the starting conformation. In contrast, the lipase shows the highest amount of structural changes in the aqueous solution.

The lipase exhibits a strong bimodal gorge radius distribution, open and closed states, for system 100H. When in the open state, the lipase yields a persistent gorge radius of 3 Å, while in the 0H system it produces narrowing gorge radius from 3 to 1 Å as water is squeezed out. In contrast, the two mixture systems (5H and 95H) actually show a remarkable widening of the gorge from 3 to 4-6 Å. The lipase in 5% hexane shows the widest maximum gorge radius of 6 Å.

We believe that the presence of minute organic moiety in a significantly aqueous environment is very desirable molecular motif for lipase activation. Two interfacial environments (5H and 95H) were created to provide some qualitative features. However, it is far from obtaining a qualitative answer to the optimum mixture and solvent structure to maximize the lipase activity. This question highlights the difficulty of studying protein at immiscible interfaces compared to that in solutions.

Properties obtained here will also facilitate the comparison of lipase conformational dynamics in other organic solvent medias and in high pressure conditions in other studies in the future. These quantitative analyses on lipase conformations and solvent structures provide a mechanism for examining the molecular immiscible interface-activated conformational switch of proteins, especially on the use of an amphiphilic helix (α5 in this case) as the critical component.

Acknowledgments

We acknowledge the computational support provided in part by the UT-ORNL Center for Molecular Biophysics, and by an allocation of advanced computing resource (TG-MCB120011) on Kraken at the National Institute for Computational Sciences.

3.6 Appendix 1

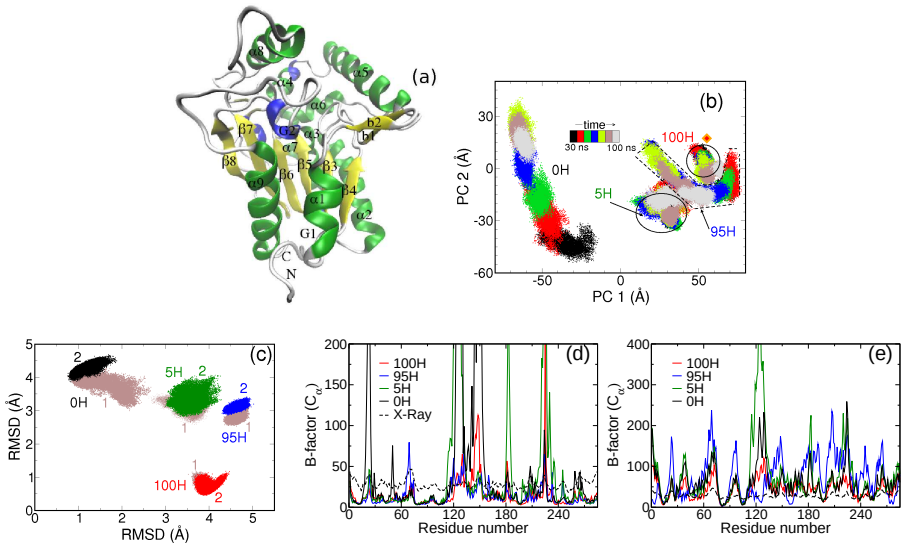

Figure 3.1: Two-dimensional PCA, RMSD, and B-factor of lipase in various solvents

(a) 1EX9 canonical view with structural elements labeled. (b) Principal component analysis (PCA) for the lipase in respective solvent conditions. PC1 and PC2 are the top two principal components (PCs), respectively. Each color block is 10 ns of data. The diamond symbol (orange) indicates the location of the original crystal conformation projected on these two PCs. (c) 2D-RMSD plot of the lipase in various solvent conditions: 100H (red), 95H (blue), 5H (green), and 0H (black). The x-axis (y-axis) indicates the RMSD of the snapshots from the mean structure of the lipase in water (hexane). Labels 1 and 2 represent the first and second half of the final 70 ns of the simulations, respectively. Cα B-factor of the lipase in various solvent conditions: 100H, 95H, 5H, 0H and x-ray (dashed line) for the last 100 ns (d) and 10 ns (e) of data.

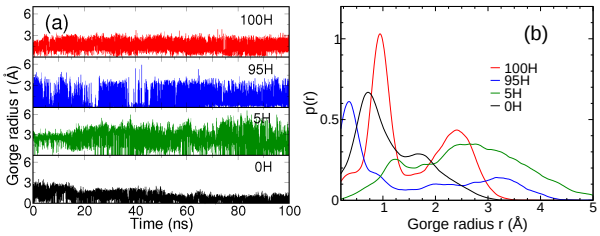

Figure 3.2: Gorge radius of lipase

The time series (a) and distribution (b) of the gorge radii of the access channel in different solvent systems. Color notation is identical to that of Fig. 1c.

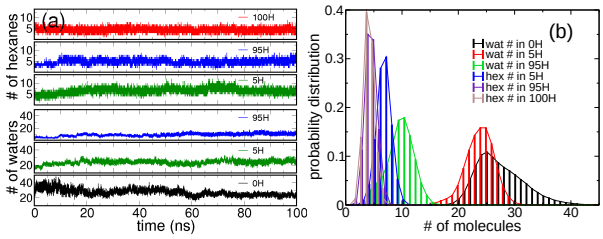

Figure 3.3: Solvent molecules residing in the gorge

The time series (a) and distribution (b) of the number of solvent molecules in the gorge.

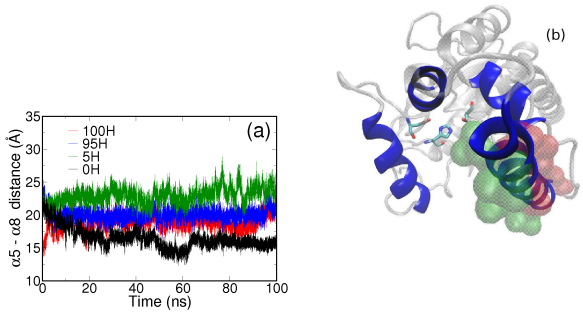

Figure 3.4: Opening of the lid is increased by oil-water interface

(a) Change in the distance of the helices α5 (Cα of Ile^{159}) and α8 (Cα of His^{215}) through time in various solvents. Color notation is identical to that of Fig. 1c. (b) A top view of the gorge entrance. Helices α4, α5, α6, and α8 are colored in blue. The three residues (stick representation) near the core of the protein are part of the catalytic triad (Ser^{82}, Asp^{229}, and His^{251}). The green and red shadings indicate the hydrophobic and hydrophilic faces of the α5 helix, respectively.

Chapter 4

Solvent-Induced α- to 3_{10}-Helix Transition of an Amphiphilic Peptide

Solvent-induced α- to 3_{10}-helix transition of an amphiphilic peptide

Ricky B. Nellas,[1,2] Quentin R. Johnson,[2,3] and Tongye Shen[1,2]*

[1]Department of Biochemistry and Cellular & Molecular Biology,

University of Tennessee, Knoxville, TN, 37996

[2]Center for Molecular Biophysics, Oak Ridge National Lab, Oak Ridge, TN, 37830

[3]UT-ORNL Graduate School of Genome Science and Technology, Knoxville, TN, 37996

Abstract

The amphiphilic peptide of the triacylglycerol lipase derived from *Pseudomonas aeruginosa* plays a critical role in guarding the gate for ligand access. Conformations of this peptide at several water-oil interfaces and in protein environments were compared using atomistic simulations with explicit solvents. In oil-containing solvents, this peptide is able to retain a folded structure. Interestingly, when the peptide is immersed in a low polarity solvent environment, it exhibits a "coalesced" helix structure, which has both α- and 3_{10}-helix components. The observation that 3_{10}-helix conformation is populated in a highly nonpolar environment is consistent with a previous report on polymethylalanine. Frequent interconversions of the secondary structure (between α-helix and 3_{10}-helix) of the peptide are also observed. We fur-

ther studied how this solvent-induced structural transition may be connected to the trigger mechanism of the lipase gating and how the lipase senses the hydrophobic-hydrophilic interface.

Amphiphilic (also known as amphipathic) peptides are surprisingly abundant in nature's design of biomolecules. [66, 67] Although hidden from the plain view of sequence information, a distinct helix containing a hydrophilic face and a hydrophobic face is revealed when the peptide is properly folded. [66, 68–70] Due to this unique amphiphilic pattern, these peptides can act as a barrier between polar and nonpolar components of a protein and can also have unique roles for proteins that function at interfaces. [66] In fact, several important peptides, from hormones, cytolytic toxins, to neurotransmitters are amphiphiles. [66, 71, 72] Also, many of the amphiphilic helices are the critical components of proteins, [66] from lung surfactant protein (water-air interface), [73, 74] to channel proteins (multi-helix packing), [75, 76] to antifreeze proteins (ice structuring) [77, 78].

Canonical helices such as the 3_{10}-, α-, and π-helix are stabilized by backbone hydrogen bonding between the carbonyl group (C=O) of residue i and the amine group (N-H) of a downstream residue $i+3$, $i+4$, and $i+5$, respectively. [79, 80] A typical α-helix has 3.6 residues per turn (100° rotation per residue) and a 5.4 Å pitch (1.5 Å per residue). The sequence information of an α-helix can be displayed in an extended wheel diagram, as shown in Figure 4.1a. Each complete layer of the wheel has 18 residues. The N-terminus belongs to the innermost layer. In contrast, a 3_{10}-helix has three residues per turn (120° rotation per residue) and a longer pitch of 5.8 to 6 Å (2 Å per residue). [81] Thus, a 3_{10}-helix displayed in an extended

wheel diagram has three residues per layer. Other noncanonical helix structures have also been suggested, such as the 11-mer repeat α11/3 helix. [82, 83]

Since the H-bonding of an α-helix needs five residues to initiate, extremely short segments are more likely to be found in the 3_{10}-helix conformation while longer helical structures are assumed to be predominantly α-helix. [84] Of the peptides that are known to adopt a helical conformation, 10% are 3_{10}-helix. [85,86] A pioneering study on the interconversion between the α- and 3_{10}-helix using a capped decamer of α-methylalanine (a nonproteinogenic amino acid) examined the thermodynamics of these two conformations in vacuum and explicit solvents such as water, acetonitrile, and dichloromethane. [87] In the crystal structures of artificial peptides that contain nonproteinogenic amino acids α-monoalkyl and α,α-dialkyl, both α- and 3_{10}-helices are commonly observed. [87] In a crystallographic examination, a section of the apo form of lactate dehydrogenase exhibits α-helical conformation. [88] Upon binding to lactate and nicotinamide adenine dinucleotide (NAD), it forms a "coalesced" helix, part of it becomes a 3_{10}-helix and the rest remains α-helical. [88] Also, the emergence of a molten helix has been observed in the open to closed transformation of aspartate aminotransferase. [89]

The general preference of α- versus 3_{10}-helix as well as the dynamics of the transition between these two states are still not well understood, especially in the context of a naturally occurring amphiphilic helix. Crystallographic data suggests

that peptide length and capping residues influence the resulting conformation of the helix in general. [90–93] Overall, longer peptides can manifest a mixture of α- and 3_{10}-helix conformations, while short peptides may favor the 3_{10}-helix structure. [94] There is a critical length of the peptide indicating the shift of the preference. Based on extensive oligo-α-methylalanine studies, the critical length at which α- and 3_{10}-helix conformers have nearly equal potential energy can be computed. The results are slightly force-field dependent which give the critical length at 12, 8, and <7 amino acids for the parameters of AMBER all-atom, [95] AMBER united-atom, [96] and OPLS, [97] respectively.

Previous experimental [81,98,99] and computational [87] studies demonstrated that the polarity of the solvent can affect the preferred conformation, α- or 3_{10}-helix. For poly-α-methylalanine, changing the solvent from dichloromethane to water increases the chance to form α- over 3_{10}-helix. [87] In addition, decamethylalanine in water or deuterated dimethyl sulfoxide favors α- over 3_{10}-helix while in a nonpolar environment such as dichloromethane or deuterated chloroform, 3_{10}-helix conformation is more stable. [100] There is a consensus that, for these simple artificial peptides, α-helix dominates in polar environment while 3_{10}-helix can be more populated in anhydrous solvents. [87, 91, 92, 98, 100] Still, little studies have been reported on how complex solvent environments (including low-polarity solvents and interfaces) affect the conformations of a natural peptide of amphiphilic

nature.

Systems and Methods

In this study, atomistic simulations were performed on amphiphilic peptides from *P. aeruginosa* lipase PAO1, and from human apolipoprotein in various explicit solvents and interfacial systems. The initial conformation of the 26-residue peptide was obtained from a lipase PAO1 (PDB:1EX9 residues 124-149, GSAGE AVLSG LVNSL GALIS FLSSG S). [41] We also examined a 26-residue peptide from a human apolipoprotein (PDB:1AV1 residues 90-115, LEEVK AKVQP YLDDF QKKWQ EEMEL Y) [101] for comparison. The force-field AMBER 99SB was used for the peptide. [102] TIP3P model [58] was used for the parameter of water while hexane and benzene were parameterized using the antechamber module of AMBER 10. [19] A sodium ion was added to neutralize each α5 system.

Various systems were created by solvating the 26-residue α5 peptide of PAO1 (hereinafter referred to as "A") in water (A100W), in a minute amount of hexane in water (A5H), at a planar hexane-water interface (A50H), a minute amount of water in hexane (A95H), and in hexane (A100H). The initial solvation procedure using multiple solvents is similar to that of Ref. [103] A complete list of the systems investigated is shown in Table S1 in the supporting information (SI). Representa-

tive snapshots of the initial and evolved configurations of interfacial systems (A5H, A50H, and A95H) are shown in SI Figure S1. The setups for the 26-residue peptide from the human apolipoprotein in water (L100W) and in hexane (L100H) are performed in a similar manner as that of peptide α5.

Minimization, heating, NPT ($p = 1$ bar) equilibration, and production runs were performed using the sander module of AMBER 10. [19] All simulation production runs were performed at 300 K and constant volume (NVT ensemble). Temperature was regulated using the Langevin thermostat. The long-range interactions were treated using the particle mesh Ewald method. A nonbonded cutoff of 8 Å and a timestep of 2 fs are used. Frames are collected at 1 ps intervals. Data collected from the 50 ns equilibration run was discarded and statistical analyses were performed on the last 150 ns of the simulations.

Additionally, data on the conformations of the peptide section from the previous simulation of the whole 285-residue enzyme [103] (hereinafter referred to as "E") will be used for comparison. As an example, the notation "E100H" refers to the system of the enzyme (E) placed in pure hexane (100H).

Results and Discussion

We will discuss below how complex solvent interfaces perturbed the structural integrity of the peptide $\alpha 5$ of *P. aeruginosa* lipase. To our knowledge, this is a first report on the solvent-induced conformational switch of a functional and naturally occurring amphiphilic peptide.

4.1 A Coalesced Helical Structure

We first report the coexistence of the α- and 3_{10}-helix conformers in peptide $\alpha 5$. As shown in Figure 4.1a (the helical wheel representation [69]), this 26-residue peptide has an amphiphilic characteristic when folded as an α-helix. The corresponding 3_{10}-helix arrangement is depicted in Figure 4.1b. One of the three branches is mainly hydrophobic. Another one is mainly hydrophilic, while the third branch contains all the glycine residues. One can also show the amphiphilicity in the non-canonical 11-mer repeat helix. [82]

To quantify the secondary structure of this helical peptide, we monitored the formation of backbone H-bonds, as illustrated in Figure 4.1c. Here, two approaches were used to determine whether an H-bond is formed for a particular configuration. Using a set of geometric criteria, an H-bond is formed when the distance ($r_{O\cdots H}$) between the carbonyl oxygen of residue i and the amide hydrogen of residue $i+4$

(or $i + 3$) is 3.5 Å or less, and the angle ($\theta_{N-H\cdots O}$) is between 120° to 180°. An alternative (energetic) criterion implemented in DSSP algorithm [104] deems an H-bond formed when the energy E between C=O of residue i and N-H of $i+4$ or $i+3$ is lower than -0.5 kcal/mol. Here E is $27.888\times(1/r_{ON}+1/r_{CH}-1/r_{OH}-1/r_{CN})$ kcal Å/mol. H-bond formation is calculated every picosecond and the mean value over the 150 ns data is reported here. Some of the hydrogen bonds are clearly bifurcated [80, 105] (C=O of residue i simultaneously forms H-bonds with N-H of both $i + 3$ and $i + 4$).

As shown in Figure 4.2a and c, the percentage of α-helix backbone H-bond in hexane (A100H) decreases gradually from N- to C-terminus, but unfolding was not observed from visual inspection. To resolve the apparent conflicting information between the loss of backbone i-$i + 4$ H-bonding and structural integrity, we examined possible i-$i + 3$ backbone H-bonds. As shown in Figure 4.2b and d, the propensity of 3_{10}-helix indeed increases from N- to C-terminus. Although it seems that the energetic criterion is more stringent than the geometric criterion both give a consistent trend. Most of our results were obtained from the geometric criterion, but the use of the energetic criterion does not change the findings.

We further studied the dynamics of backbone H-bond formation. H-bonding using geometric criterion was calculated and the mean value from a running nanosecond window is presented here. The time evolution of H-bond (shown in Figure 4.2e)

demonstrates that the peptide contains both α- and 3_{10}-helix conformations. While the total number of backbone H-bonds is conserved approximately, there is a switch between the dominant type of H-bonds at the order of tens of nanoseconds. Since the secondary structural elements of the peptides interconvert between α- and 3_{10}-helical conformations, we would like to further examine this switch in a more site-resolved resolution. We performed similar analysis on the dynamics of backbone H-bond formation by obtaining separate statistics on the N- and C-terminus halves of the peptide (SI Figure S2). Interestingly, when the N-terminus half is largely α-helix, the C-terminus half is 3_{10}-helix and vice versa. On average, the N-terminus is rich in α-helix conformers while the C-terminus is mostly 3_{10}-helical. The existence of this dynamic switch at the level of secondary structure can be corroborated from the dynamics of the root-mean-square deviations from the initial α helical structure (SI Figure S3). The dynamic fluctuation of RMSD strongly correlates with the dynamics of hydrogen bonding. Connected with the dynamic H-bonding analysis, it seems that the RMSD values below 4 Å pertain to the α-helix-rich structure while values above 4 Å suggest a conformation rich in 3_{10}-helix secondary structures.

Besides hydrogen bonding and RMSD, the Ramachandran plot (backbone torsional angles ϕ-ψ) is often used to characterize the secondary structure. The coalesced helical structure for A100H is also evident from the the Ramachandran plots

shown in Figure 4.3. Though both are at helical regions, the locations of the free energy minima in A50H ($-65°, -40°$) and A100H ($-60°, -25°$) are dramatically different. An important ϕ-ψ signature that distinguishes different helical structures is the value of ϕ + ψ. The population of typical α-helical structures peaks around $\phi + \psi = -104°$ while 3_{10}-helix gives $-75°$. [106]. Here A100H gives $\phi + \psi = -105°$, which is consistent with the typical value of the α helix, while A50H gives $-85°$, a value that is between the α and 3_{10} limits. Thus, the population for the A100H system is shifted to the upper right direction which signals an enhanced sampling of the 3_{10}-helix conformation. Using these complementary analyses, our result clearly show a "coalesced" helix with sub-regions of different types of helices that interconvert.

4.2 Conformations of Amphiphilic Peptides in Complex Solvent Milieux

Besides the study of the conformations of α5 in hexane, which is reported in the previous subsection, we also studied α5 at different hydrophobic-hydrophilic interfaces. When the α5 peptide is in a significantly nonpolar environment (A100H, A95H, or A50H), α5 has an α-helix structure, as shown in Figure 4.4a-c (using the geometric criterion) and Figure 4.3. Meanwhile, the end-to-end distance (the

distance between the backbone nitrogen on both ends, 1N-26N) shows little fluctuation. As shown in Figure 4.5, the mean end-to-end distance is about 42 Å for A100H and 37 Å for A50H. A much shorter mean end-to-end distance of A95H ($\sim$8 Å) is caused by a predominant helix-turn-helix conformation. As indicated in Figure 4.4b, there is a break of the helix around the midsection. By visual inspection, it was observed that water molecules clamp the middle of the peptide, around position 14 (Ser137), which is a hydrophilic residue located at the center of the hydrophobic face.

On the other hand, when the α5 peptide is in a significantly polar environment (A5H and A100W), it is a disordered peptide with a diverse set of ϕ-ψ values (Figure 4.3). In this polar condition, most of its residues do not participate in backbone H-bonding (Figure 4.4d-e) and its structure is highly perturbed with broad end-to-end distance distribution spanning $\sim$5 to 40 Å. This shows that the presence of water molecules plays an important role in the conformational exploration of peptides. [63, 107] The α-helix disruption in this system can be attributed to the presence of a stronger intermolecular (peptide-water) than intramolecular (intrapeptide) interaction.

So far we found that the amphiphilic peptide, α5 of lipase PAO1, is stable in pure hexane and at a planar hexane-water interface but not in pure water. It is important to see how general the conclusion is. Besides the peptide α5 of PAO1,

we also studied the peptide $\alpha 5$ of extremophilic strain of *P. aeruginosa* (PALLST03), [56, 57] Am100H, which has only one different residue, V130I. Not surprisingly, the change from valine to isoleucine did not change the nature of helix conformation reported so far. However, not all amphiphilic helices share the stability of $\alpha 5$ in anhydrous solvents. Our results from L100H indicate that, unlike $\alpha 5$, the selected peptide from human apolipoprotein mostly unfolds in pure hexane. However, sections of the peptide, residue 3-6 (α-helix) and 13-20 (3_{10}-helix), still display a strong secondary structure. While all the amphiphilic peptides examined in this study are stable at a planar water-oil interface. On the average, the amphiphilic peptides examined here are unfolded in water. Interestingly, residues 3-6 and 9-12 of the human apolipoprotein participate in the α-helix formation while residues 16-19 are 3_{10}-helical in water. This indicates that the peptide composition and sequence (side-chain identity) may be crucial. Also, it seems that the presence of both α-helix and 3_{10}-helix conformations is ubiquitous in proteinogenic amphiphilic peptides. Note that the examined section of the apolipoprotein has a higher percentage of hydrophilic residue than that of $\alpha 5$.

4.3 Interfacial Water

To understand the role of interfacial water on the conformations of the peptide, we applied a solvent mapping procedure to locate the peptide hydration sites, a region around the peptide where water molecules preferentially reside. [108] For the systems that stay α-helical, we can further express the density of water in terms of cylindrical coordinates (d, ϕ, z) using an ideal α-helical reference state (created by the LEAP program). Here d is the radial coordinate, ϕ is azimuth and h is height. We set the coordinates by aligning the direction of the ideal helix as the cylindrical axis $\hat{\mathbf{z}}$ and set the C_α of the first residue, Gly, at the direction of azimuth $\phi = 0$. After the integration along z, we express the water density $\rho(\phi, d)$. We applied this analysis to the A50H system. This integration procedure is appropriate only if the peptide is not dramatically different than that of the ideal helix reference. Generally we can always express the water density in three-dimensional Cartesian coordinates.

Water densities surrounding peptide α5 in A5H, A50H, and A95H systems are shown in Figure 4.6. For A5H, the presence of water molecules in most regions of the peptide results in a lower backbone α-helical H-bond formation and an enhanced conformational exploration (Figure 4.3b) of the peptide as compared to that of A50H. For A95H, only the termini and mid-section of the peptide are highly hydrated which leads to broken mid- and terminal backbone H-bonds and thus favors

a helix-turn-helix conformation. This preferential hydration in A95H gives rise to the reduction of the conformational exploration of the peptide (Figure 4.3b) and the retention of many of its backbone H-bonds (Figure 4.4b).

For A50H, the α-helical peptide stays at an almost planar interface. As shown by $\rho(\phi, d)$ in Figure 4.6d, the region ($-160^\circ < \phi < 60^\circ$) has low water density. This is the region where the hydrophobic residues are located, as shown in the wheel diagram (Figure 4.1a). This solvent structure in the A50H system thus stabilize this amphiphilic peptide, which has limited conformational exploration (Figure 4.3c) and high α-helical H-bonding (Figure 4.4c).

4.4 Solvent Specificity

Previous evidence shows activity preference in the pairing of specific lipases and organic solvent types (aromatic vs. aliphatic). [10, 32] For example, the activity of PAL-LST03 [32] monotonously increases with the ratio of methyl:phenyl (from benzene, toluene, to p-xylene), while *A. baylyi* [10] shows the opposite trend. Motivated by these reports, we also examined the specificity of the anhydrous solvents on the conformations of peptide α5. By monitoring the percentage of backbone H-bond (α- and 3_{10}-type) of the α5 peptide in hexane (A100H) and in benzene (A100B), we compared the helical propensity in aliphatic and aromatic solvents.

While the peptide in hexane exhibits both α- and 3_{10}-helix conformations, the peptide in benzene is mostly α-helical, as indicated by H-bonding (Figure 4.4f). The end-to-end distance results also confirmed that the peptide in hexane is more elongated than when it is in benzene with mean distances of $\sim$42 and 35 Å (Figure 4.5), respectively. Additional studies on α5 peptide at interfaces, water-hexane (A50H) and water-benzene (A50B), showed both are rich in α-helix conformers. As shown in Figure 4.5, A50H is more elongated than A50B. The terminal H-bonds are broken in A50H while they remain intact in A50B (Figure 4.4c).

Here we did not observe the transition from the initial α-helix to a 3_{10}-helix-rich conformation for A100B, rather, α-helix shows strong stability. It is possible that there is a kinetic factor that makes benzene a strong immobilizing media. The viscosity of benzene (0.59 cp) is higher than that of hexane (0.29 cp). [109] The potential transition from the initial α-helix to 3_{10}-helix-rich structure may still occur, but in a longer time than the sub-microsecond timescale reported here. The absence of 3_{10}-helix in benzene may also have an equilibrium factor: the larger relative polarity of benzene (0.111) compared to that of hexane (0.009). [110] At the atomistic level, carbon and hydrogen atoms of benzene have larger absolute partial charges than their hexane counterparts.

4.5 Enzyme Environment

Finally, we report the structure of the $\alpha 5$ section of the lipase from the previous whole protein simulations (100 ns). [103] As shown in Figure 4.4, regardless of solvent conditions, the backbone H-bonds tend to be more stable when the peptide is a part of the enzyme than the stand-alone peptide. A localization of the ends and a decrease of solvent exposure due to the presence of the rest of the lipase may contribute to the enhanced stability. The $\alpha 5$ section of the enzyme exhibits the coexistence of α- and 3_{10}-helix for all four solvent conditions. Here, the calculated mean number, N of i-$i+4$ (i-$i+3$) H-bond values are N(E100W) = 11.5 (5.3), N(E5H) = 12.8 (4.9), N(E95H) = 16.0 (6.0), and N(E100H) = 13.7 (6.4).

It is interesting to speculate the importance of the coexistence of α- and 3_{10}-helix-rich structure to the function of the lipase gating. One possibility is that the transition from α- to 3_{10}-helix is the environmental sensing or trigger mechanism of gate opening. Another possibility is that α- and 3_{10}-helix are stable and predominant at different solvent conditions and thus enhance the overall stability of the folded helical conformation. We performed further analysis to examine whether there is any simple correlation between the level of gate opening measured by gorge radius [103] and the conformation of $\alpha 5$ in four solvent environments: E100W, E5H, E95H, and E100H. The mean gorge radius values are r(E100W) = 0.7, r(E5H) = 1.4, r(E95H) = 2.3, and r(E100H) = 1.3 Å. Thus, it seems that gate opening is pro-

moted when the helix (either in α- and 3_{10}-helix form) is stable. Thus, a structural buffering mechanism to stabilize the helical structure might be essential to keep the gate open.

Conclusions

In summary, this work examines the solvent-induced conformational preference in a naturally occurring amphiphilic peptide from *P. aeruginosa* lipase. Here we have compared the behavior of the peptide from the perspectives of peptide chemical identity, solvent interface composition, solvent specificity and enzyme environment. We found that the peptide can exhibit an interesting transition between α- and 3_{10}-helix-rich conformations which is consistent with previous studies on the stabilizing effect of 3_{10}-helix conformation by low polarity solvents. At water-oil interfaces, the retention or disruption of H-bonds at a specific site is caused by the preferential distribution of water molecules around the peptide, and essentially, the specific interaction of the residues with water.

The peptide behaves differently when it is a part of the enzyme. Interestingly, the $\alpha 5$ section of the enzyme shows molten helix characteristics. We further discussed two possible scenarios (gating trigger mechanism vs conformational buffering) of how this partial 3_{10}-helix feature may be important for lipase function. We

concluded that helical interconversion is a structural buffering act in a variety of solvent environment to keep the access channel open.

Acknowledgments

We acknowledge the computational support provided by the UT-ORNL Center for Molecular Biophysics, and by allocations of advanced computing resources (TG-MCB120011, TG-MCA08X032, UT-TENN0049) on Kraken at the National Institute for Computational Sciences.

4.6 Appendix 2

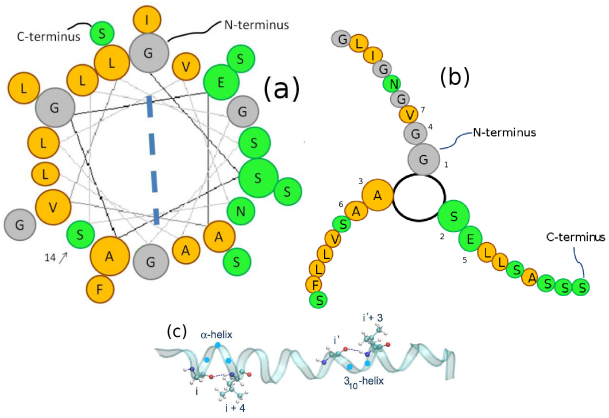

Figure 4.1: Helical wheel cartoon

(a) An α- and (b) 3_{10}-helical wheel representation of the 26-residue peptide from *Pseudomonas aeruginosa* lipase. For the α-helical wheel, each layer of the wheel has 18 residues. For the 3_{10}-helical wheel, each layer has three residues with the first few residues labeled. This visualization shows that this naturally occurring peptide has an amphiphilic character, wherein the hydrophilic residues (green) are on one side of the wheel while the hydrophobic residues (orange) are on the other side of the wheel. (c) Representative snapshot of the 26-residue peptide exhibiting both α- (i-i+4) and 3_{10}-helix (i-i+3) conformations. Circles in the ribbon represent Cαs.

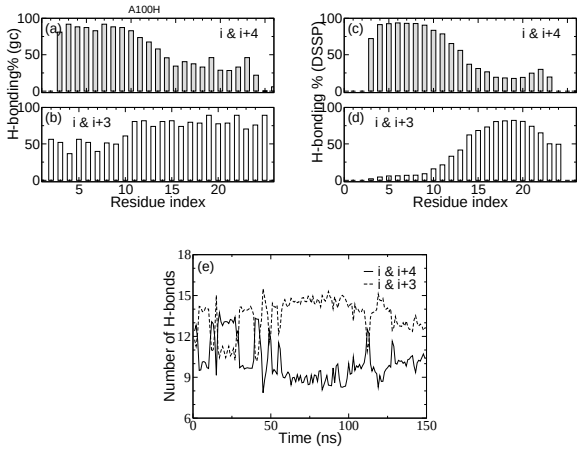

Figure 4.2: Dynamic hydrogen bond percentage

Percentage of backbone H-bonds made between residues i (CO) and $i+4$ (NH) of the α-helix conformation and i (CO) and $i+3$ (NH) of the 3_{10}-helix conformation in A100H. (a-b) Geometric criterion: An H-bond is formed when the distance ($r_{O\cdots H}$) between the carbonyl oxygen of residue i and the amide hydrogen of residue $i+4$ (or $i+3$) is 3.5 Å or less, and the carbonyl oxygen of residue i and the amide NH of residue $i+4$ (or $i+3$) forms an angle ($\theta_{N-H\cdots O}$) of around 120° to 180°. (c-d) H-bonding using DSSP energetic criterion. H-bond formation is calculated every picosecond and averaged over 150 ns of simulation data. (e) Time evolution of the total number of H-bonds formed (α- and 3_{10}-helix conformation) in the peptide for A100H.

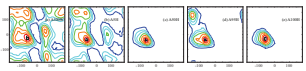

Figure 4.3: Ramachandran plot of various systems

The free energy profiles of Ramachandran for (a) A100W, (b) A5H, (c) A50H, (d) A95H, and (e) A100H at 300 K. The Ramachandrans of all the residues of the peptide (except the ones at termini) were used to construct each profile.

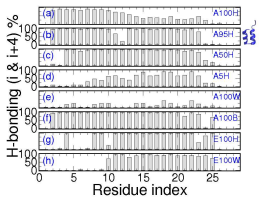

Figure 4.4: Average hydrogen bond percentage

Percentage of backbone H-bonds made between residues i (CO) and $i + 4$ (NH) constituting the α-helix conformation in various solvent milieux. In all systems, H-bond formation (geometric criterion) is calculated every picosecond. Mean values are obtained from 50 ns data in E100H and E100W (previous study [103]) and 150 ns in other systems.

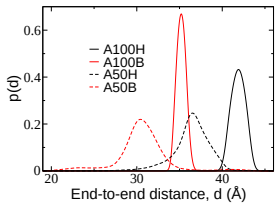

Figure 4.5: End-to-end of various peptides

The distribution of the end-to-end distance between the backbone nitrogen on both ends (1N-26N) of the 26-residue peptide α5 in different solvent conditions.

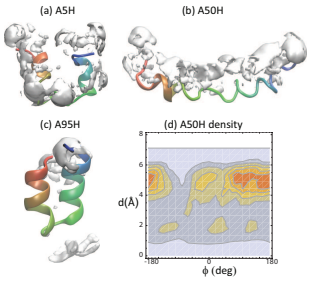

Figure 4.6: Water density around various peptides

Water densities surrounding the peptide α5 in the A5H, A50H, and A95H systems are displayed in (a-c) as iso-density contours, respectively. For each system, the structure of the peptide at the end of a 200 ns simulation was used as the reference state in calculating water density around the peptide. (d) Density of water molecules of A50H expressed in the cylindrical coordinates, $\rho(\phi, d)$. Color red indicates high density of water.

Chapter 5

Pressure Induced Conformational Switch of an Interfacial Protein

Pressure induced conformational switch of interfacial protein Lipase

Quentin R. Johnson[1,2], Richard J. Lindsay[2,3], Ricky B. Nellas[4] and Tongye Shen[2,3]

[1]UT-ORNL Graduate School of Genome Science and Technology, Knoxville, TN, 37996

[2]Center for Molecular Biophysics, Oak Ridge National Lab, Oak Ridge, TN, 37830

[3]Department of Biochemistry and Cellular & Molecular Biology,

University of Tennessee, Knoxville, TN, 37996

[4]Institute of Chemistry, University of the Philippines Diliman, Quezon City, Philippines

Abstract

A special class of proteins adopts an inactive conformation in aqueous solution and activates at an interface (such as the surface of lipid droplet) by switching their conformations. Lipase, an essential enzyme for breaking down lipids, serves as a model system for studying such interfacial proteins. The underlying conformational switch of lipase induced by solvent condition is achieved through changing the status of the gated substrate-access channel. Interestingly, a lipase was also reported to exhibit pressure activation, which indicates it is drastically active at high hydrostatic pressure. To unravel the molecular mechanism of this unusual phenomenon, we examined the structural changes induced by high hydrostatic pressures (up to

1500 MPa) using molecular dynamics simulations. By monitoring the degree of access-channel openness, we found that the protein undergoes a conformational transition and opens the access channel at high pressures ($>$100 MPa). Particularly, the amphiphilic $\alpha 5$ region of the protein becomes more ordered in response to increasing pressure. This positive correlation between the open channel and ordered $\alpha 5$ is consistent with the early finding of the gating motion in the presence of a water-oil interface. Statistical analysis of the conformations also reveals the essential collective motions of the protein and how these motions contribute to gating. However, at extremely high pressures ($>$750 MPa), the ordered $\alpha 5$ is still observed in simulations while the value of the gorge radius is low, and it seems the protein is transitioned into a different, dynamically trapped state. Arguments are presented as to why heightened sensitivity to high pressure perturbation can be a general feature of switchable interaction proteins. Further mutations are also suggested to validate our observation.

Understanding how proteins change their configurations in response to environmental perturbations can be essential for decoding how they achieve biological functions. This is especially true for a class of "switchable" proteins that have multiple populated conformational states. Lipase is one such important protein which is activated at a water-oil immiscible interface. [10, 30, 33] It can serve as a model system to study a class of important yet unexplored interfacial proteins that function at biological interfaces (such as the surface of lipid droplets). [111]

Triacylglycerol lipases catalyze the hydrolysis of glycerol esters. [34, 35] These interfacial enzymes are not only important for many biological functions, but also essential in many industrial applications. [36] As a member of the α/β-hydrolase superfamily, [34, 37] lipases exhibit the canonical catalytic triad. The mechanism at the catalytic site is deemed to be rather efficient and once the substrate reaches the active site. [34] Researchers have compared the "open" and "closed" crystal structures of lipases. [112] Therefore, the subject of lipase gating (how the enzyme regulates ligand admission to the active site) can be a focal point of lipase enzymatics. Triacylglycerol lipases, such as those from *P. aeruginosa* and *C. antarctica*, generally have a buried active (catalytic) site which substrates are required to enter through a narrow channel (gorge). Admission to the active site is controlled by a largely hydrophobic channel. [41] A particularly important area of the channel is constructed by several helices which have been referred to as the "lid" region of

the lipase. It has been suggested that movement of the lid is essential for ligand admission and, thus, enzyme activity. [37, 38, 41, 112] Previous simulation studies have revealed the molecular details of the conformational switch from a closed-gate state in aqueous solution to an open-gate state at interface. [103, 113]

More intriguingly, it seems that the close-to-open conformational switch can be triggered by high pressure as well. A series of pioneering experiments on lipase enzymatics under high pressure were reported previously, [114, 115] which found that the lipase *C. antarctica* B (CALB) shows heightened activity with increasing pressure (10 to 400 MPa). [116, 117] This phenomenon was thought to be related to the pressure-induced conformational changes of the lipase, but post-incubation effects, such as immobilization support and depressurization, could not be ruled out from the experimental designs. [116, 117] We think that it is quite plausible that such enhanced activity is connected to the gating dynamics of lipases, as postulated in Figure 1. Thus, it is interesting to investigate possible origins of structural and dynamic changes [107] in high pressure environments from a computational perspective. In the current study, we aim to probe the effect of pressure on lipase structure using all-atom molecular dynamics (MD) simulation.

Indeed, high hydrostatic pressure can have profound effects on the configurations of proteins. Pressure generally has a destabilizing effect on protein structures. It is well known that extreme pressures can induce instability and even

protein unfolding. [118–120] Increasing pressure p also generally decreases enzymatic activity, which can be characterized by an Arrhenius rate equation, $k \simeq \exp(-p\Delta V/k_B T)$ with a positive activation volume ΔV. [121, 122] However, the lipase reported in the experiments has an unusual and intriguing negative activation volume. [116] There are several benefits (biological, industrial, and academic aspects) to the study of high pressure effects on the conformations of biopolymers. One straightforward reason is to further our understanding of naturally occurring barophilic species (extremophiles that thrive under high pressure conditions). [121] These species can adapt to high pressure not only at the genetic and systems biology level, but also at molecular structural level. Beyond the extreme hydrostatic pressure (up to 120 MPa) that occur in volcanic sea vents, man-made conditions can reach much higher pressures. Thus, another benefit comes from industrial applications, such as in the food industry. [117] Often, biochemical processes conducted in an artificial setting use high pressure for desired results. Usual industrial applications of high pressure are for deactivating deleterious enzymes. However, there is experimental evidence that pressure can increase activity of a small subset of enzymes, such as lipases and acetylcholinesterase. [117] Besides practical applications, there are additional benefits for academic pursuit. Pressure provides a unique biophysical probe of protein-solvent interactions (especially the hydrophobic, hydrogen-bonding, and hydration aspects) and can provide a fundamental un-

derstanding of the mechanics of protein stability. In general, high pressure increases the importance of polar interactions between protein and aqueous solution (promoting a more extended structure), and weakens the strength of hydrophobic packing (promoting a compact structure) at the same time. This shift can often changes the balance in terms of thermodynamics and thus the preferred conformations.

In this study, we report enhanced gating motions and opening of the access channel due to high pressure. We attempt to connect the gating motions with detailed structural changes of the protein under pressure. We found that one important helical region, α5, increases its structural ordering with increasing pressure which may be responsible for the gating motion. Besides traditional analysis of snapshots from atomistic molecular dynamics, we also performed a coarse-grained analysis to locate the collective conformational changes induced by pressure. These methods can be useful for the study of proteins under high pressure in general.

System and methods

A series of atomistic simulations were performed on lipase in an aqueous solution at 300 K and pressures of 100, 250, 500, 750, 1000, and 1500 MPa. We selected a lipase from *P. aeruginosa* PAO1 for this study, since it was the subject of our early studies of lipase at water-oil interfaces. This facilitates comparison of the

conformational changes induced by pressure with the previous solvent perturbations. Also, PAO1 lipase shares the same fold as the lipase from *C. antarctica* (PDB:1TCA). Simulation data for the PAO1 lipase at room pressure was obtained from a previous work which examined lipase gating motions at an immiscible solvent interface. [103] The initial structure of the protein was obtained from crystal structure PDB:1EX9 (with ligand removed). [41] The protein was placed into a periodic box and solvated with 10,465 water molecules. Finally, three sodium ions were added to neutralize the system. In addition, a tightly bound calcium ion was reported be important to the structural integrity of this lipase. However, the metallic bonding properties of the calcium can be quite challenging to treat efficiently, therefore, we opted to fix the position of the calcium atom and the four coordinated oxygens as previously described. [103] All simulations performed under different pressures have the same starting configuration.

The AMBER 99SB force-field [102] and the TIP3P water model [58] were used for the protein and water, respectively. Pre-production minimization, heating, and equilibration (for approximately 1 ns) runs were conducted and followed by a 150 ns MD production run using the NAMD2.7 package. [59] Temperature (300 K) and the respective simulation pressure (100, 250, 500, 750, 1000, and 1500 MPa) were kept constant throughout (NPT ensemble). Long-range interactions were treated using the particle mesh Ewald method, [123] all bonds involving

hydrogen atoms were constrained using the SHAKE algorithm, [24] and pressure control was achieved using the Berendsen method. Other pertinent simulation details are similar to those reported previously. [103] Each simulation was conducted with a 2 fs time step for 150 ns, and snapshots were generated at 1 ps interval, producing 150,000 snapshots for each system besides the trajectories obtained previously at room pressure. [103]

Collective motion of the lipase gorge regulates ligand entry to the active site and, thus, is essential to the catalytic activity of the enzyme. An analysis of this ligand-access channel and its dynamics is needed to characterize how pressure may affect lipase gating. The gorge radius is defined as the maximum radius of a spherical ligand capable of passing through the channel and reaching the active site. The active site patch is constructed by Ser^{82}, Tyr^{27}, HIE^{14}, HIE^{81}, Met^{16}, Leu^{12}, Ala^{13}, Ala^{115}, Gly^{106}, and/or Leu^{231}. The program HOLE, [60] employed in the study of ion channels and protein cavities, [60–62, 103] was used here to measure the pore radii along the channel with the minimum radius reported as the gorge radius. This analysis effectively identifies the size of the bottleneck that gates ligand entry. Analysis was performed every 10 ps. Definitions of user-specified parameters are identical to those in the previous report. [103]

We also characterized the lipase motions at the level of residue-residue contacts. This approach may reveal the gating mechanism and suggest future biomolecu-

lar engineering possibilities that may alter the gating properties. Describing high-pressure induced conformational changes at the level of residue-residue contacts can be challenging in comparison to other perturbations, such as denaturants, where the effect of eliminating (native) contacts is more visible. In the endeavor of identifying important residue-residue contacts, a coarse-grained analysis method has been developed in a previous study which decoded the negative allosteric mechanism of the RXR:TR nuclear receptor complex. [124] The details of this method, Computation of Allosteric Mechanism by Residue-Residue Associations (CAMERRA) have been expounded upon in a previous publication. [124] This method involves the statistical analysis of residue-residue contacts. First, the contacts that are not always formed or never formed, which we termed "dynamic" contacts are selected. Secondly, we performed the principal component analysis (PCA) on these dynamic contacts. The end product is the collective motions expressed as "vibrational" modes of contact degrees of freedom.

Results and discussion

The access channel opens at high pressure

As stated previously, substrate access is regulated by a narrow channel that connects the buried active site to the surface of the protein. We have employed the HOLE

program to measure the size of this access channel throughout the simulation, which we refer to as the gorge radius. This effectively quantifies the maximum size of a spherical ligand that can traverse this pathway to the active site. Fig. 5.2a shows the dynamics of the gorge radius for each pressure for the entire 150 ns trajectory.

In a previous study, it was found that an aqueous environment at room pressure (1 atm) causes the channel to contract from the initial open state and results in a closed-gate conformation. [103] Meanwhile, an aqueous-oil interface can dramatically affect lipase gating by inducing an open-gate state. The gorge radius at room pressure (1 atm) has been shown in Fig. 5.2a as a point of reference. In the current study, we find that pressure also has a pronounced effect on lipase gating. One would expect that high pressure would compress the protein and thus keep the gate closed. Counter-intuitively, our simulations reveals that the protein adopts an open gate conformational ensemble at high pressures, as shown in Fig. 5.2. Meanwhile, we observe that the level of fluctuations decreases for systems at extremely high pressure (750 MPa and greater), which seems to suggest that the conformational exploration of the protein has been greatly restricted at those conditions. It is possible that high pressure introduces certain "glassy" dynamics to the protein system, thus lipase may stuck at a partially-open-state for most of the simulation.

So far, we have provided a general picture of how high pressure triggers the gating motion of the lipase. We seek to further quantify the statistics (open to

closed ratio, Fig. 5.2b) and dynamics (switching frequency, Fig. 5.2c) of gating, respectively. We utilize a cutoff value for gorge radius to differentiate between the conformations of lipase that are in either an open or a closed state. Since the exact size of the ligand is difficult to determine and approximations have been made to evaluate the level of gate opening by gorge radius, we present the results as a function of a variable cutoff.

The ratio between open and closed states as a function of a specific gorge-radius cutoff is displayed in Fig. 5.2b. The ratio of open to closed-state conformations is dependent on the gorge radius cutoff as the system is deemed less likely to adopt an open conformation as the criterion becomes more stringent. Fig. 5.2c shows how often the lipase switches between an open and closed-state conformation. Upon comparing Fig. 5.2b and Fig. 5.2c, we note that in the high cutoff region these two sets of curves are quite similar, which may be caused by the short dwell time of the lipase being in an open state. Also, it makes sense that the cutoff value (abscissa) which yields an ensemble with 50% open-state conformation (where ordinate equals 0.5 in Fig. 5.2b) is roughly corresponding cutoff that maximizes switching frequency (maximum values of ordinate reached in Fig. 5.2c).

The analysis of switching frequency is important for studying gated ligand access for various important proteins. Two essential time scales are often compared to distinguish different scenarios of gated diffusion-limited reactions. One important

time scale is the time required for the ligand to diffuse across the channel while the other is the time scale of the "gate" changes between open and close state, $1/f$ as we have shown above. The application of the theory to the practical values of lipase properties obtained from simulation suggests that lipase is operating at the "fast gating" limit, i.e., the former is a lot longer than the latter. Thus, the turnover number of lipase is not being fractioned by the percentage of gate being open, but rather a very small fraction of open state will lead to lipase being functional.

5.1 Structural ordering of α5 under high pressure

In a previous work on solvent-mediated lipase gating, it was demonstrated that gating of this enzyme can be mitigated by maintaining the structural ordering of the α5 helix. [113] This helix is part of the lid region that regulates ligand entry into the gorge which leads to the active site. It was determined that α5 could assume either an α helical or a 3_{10} helical conformation. But regardless of the type of helix, the structural ordering of α5 is positively correlated to the gorge radius. That study was conducted under room pressure where the ordering of α5 was purely affected by the solvent condition. It is therefore interesting to examine the structure of this important helix under high pressure in the current study. We want to address whether this helix is able to maintain its order since high pressure typically destabi-

lizes intra-protein hydrogen bonding by strengthening hydrogen bonds between the solvent and protein.

To determine how pressure affects the ordering of this region, we monitored the formation of backbone hydrogen bonds. Fig. 5.3 depicts the hydrogen bonding ratio (between residue i and $i+4$) in α5. Interestingly, high pressure increases the helical propensity of α5, especially on the N-terminal half of the helix. We suspect that this is due to the amphipathic nature of the helix. As α5 is partially hydrophobic, it does not form a helix in water normally, but instead a disordered structure with a packed hydrophobic core, as previously observed. [113] High pressure may alter this packing and promote the formation of a helix. The ordering of α5 may indicate that pressure-mediated gating operates by a similar mechanism as that of solvent-mediated gating. [113] It seems that this region plays an important role in high pressure gating as we will continue to examine in the following sections where the global structural changes are being examined.

5.2 Pressure induced collective motions of the lipase

In earlier sections, we condensed atomistic protein conformations down to a single value that is indicative of the size of the access channel known as the gorge radius. Here, we aim to characterize the underlying protein conformations at the

residue-residue contact level. This data reduction analysis has been utilized to evaluate processes such as protein folding, conformational switch, and allostery in the past. [124–126]

The first step of data reduction protocol is locating dynamic contacts. We sort each contact by the contact percentage during the simulation, from 100 to 0%. The resulting plot is termed a ranked contact curve (RCC). Fig. 6.2 contains a ranked contact curve for each system. This figure also contains a collective RCC that is created by combining contact information from all systems. We filtered the less dynamic contact information and concentrate on "dynamic" residue-residue contacts that are formed and broken during the course of the simulation. [124] Those contacts that are either always formed, or almost always formed, and the contacts that are either never formed, or almost never formed, do not contribute much to the motions of the protein. We set the cutoffs at 80% for those formed and at 20% for rarely formed contacts. Any contacts that are formed for 20-80% are deemed "dynamic" while 0-20% or 80-100% are "static" and will be excluded from further analysis. The number of dynamic contacts (J) is $J \equiv J_{20} - J_{80}$. Fig. 6.2 provides the combined RCC which shows that $J_{20} = 1975$ and $J_{80} = 1659$, therefore $J = 316$ in contrast to $N \times (N-1)/2 = 40470$ potentially possible contacts for a protein having $N = 285$ residues. This method may help us to isolate the important contacts that contribute to the collective gating motions of the enzyme.

As shown in Fig. 6.2, RCCs of different pressures are distinct. The RCC can also be viewed as a measurement of compactness of the system. Fig. 6.2 shows that RCCs move to the right (and up) with increasing pressure, since a packed structure has more contacts formed. In spite of high pressure opening the gorge, the overall protein has indeed become more compressed.

We further inquire on the important collective motions encoded by these "dynamic" contacts. A statistical analysis is performed on the selected $J = 316$ contact degrees of freedom (DOFs). [125] Principal Component Analysis (PCA) is a well-known linear transformation scheme to identify the collective mode of protein motion. In the current case, the covariance matrix used by PCA describes the level of correlation between the contact events. Therefore, the eigenvectors (the so-called principal components) obtained from this analysis provide the major "vibrational" modes of protein dynamics in terms of breaking and forming of residue-residue contacts. In an ideal situation, the largest few PCs dominate the fluctuation of contacts and the corresponding eigenvalue Λ_i will drop quickly with index i. In the current case, the top five eigenvectors contribute to 30% of the protein motions.

Fig. 5.5a-b are eigenvectors (PC1 and PC2, respectively) color-coded on the residue-residue contact map. Here, a positive (blue) value of the component of the eigenvectors indicates that a particular contact is being formed, and a negative (red) value indicates that a contact is breaking. PC1 (Fig. 5.5a) shows largely concerted

contact dynamics, as most contacts are forming along this PC and yields a plot dominated by positive values. This PC is largely the mode of protein motion that is linked to compression, i.e. the protein becomes more compact as more contacts are formed. PC2 (Fig. 5.5b) shows a more balanced contact forming and breaking.

Fig. 5.5c-d express the same eigenvector information as that of a-b displayed on the crystal structure of the lipase. Each cylinder in Fig. 5.5c-d indicates a contacts between two residues (connecting the Cα atoms) and depicts the corresponding component of the PC value for the contact of that residue pair. Blue indicates contact are being formed and red contact breaking. This figure assists in visualizing how the dynamic contacts contributed to the collective modes. For PC1, most of the movement of contacts are shown in the interior of the protein. Again, this may be an indication that PC1 is representative of protein compression produced by the high pressure systems. In contrast, PC2 shows more localized motions since Fig. 5.5d has local patches of red and blue regions. For the critical α5 region, PC2 indicates that the intra-helical bonds are formed. Meanwhile, the inter-helical contacts between α5 and α4 are breaking, which is indicative of the gate opening. This may suggest that PC2 is linked to the gate opening motion of the enzyme.

Finally, all simulation snapshots of each system are projected onto PC1 and PC2 in Fig. 5.6 so that each snapshot of the protein is condensed to a single point. Again, it seems that the major mode, PC1, has picked out the compression mo-

tions that stem from increasing pressure. As pressure is increased to 750 MPa and beyond, each ensemble shifts to the right along the PC1 axis relative to the room pressure system (1 atm). It seems that PC2 can be correlated to the gorge opening at room to moderately high pressure (250 MPa), i.e., high PC2 values indicate an open ensemble for these systems. It is however complicated at higher pressures. The highest pressure systems have positive PC2 values but a closed gate. These systems have a large movement in PC1 direction in comparison with the remaining system. Corroborated with the hydrogen bonding analysis of the $\alpha 5$ region, we suspect that the protein conformations have dramatic changes at these high pressures, so that an ordered $\alpha 5$ region is no longer a good indicator for the gate to be open, as other features becomes the limiting factor.

Pressure effects on switchable interfacial proteins

To put the current results of lipase conformational changes under high pressure into a general context, we have made comparisons with the early report of lipase at interface [103, 113] throughout this report. We observed that pressured induced gating shares similar characterizations as interface induced gating, especially in terms of motion at the lid region. Switchable interfacial proteins use unique amphiphilic solvent conditions to their advantage in general. Since one biophysical origin of

pressure perturbation is changing how protein and polar solvent interact, we suggest that pressure activation of switchable interfacial proteins can be a general theme.

To further make predictions to test our hypothesis on the lipase, we would suggest mutagen sis experiments on lipases to enhance or decrease lipase activities at high pressure. Since the simulation data suggests an ordered first half of the amphiphilic $\alpha 5$ region is the key, we suggest mutation of Gly^{127} and Gly^{133} to hydrophobic residues such as Leu. These two glycines are on the hydrophobic side of amphiphilic helix, thus these mutations may enhance the helicity of $\alpha 5$ and promote gating motion.

Concluding Remarks

A study of the effects of high pressure on enzyme gating dynamics using molecular dynamics simulation and computational analysis techniques is presented here. Moderately high pressure induces the lipase of *P. aeruginosa* PAO1 to adopt an ensemble of open-gate configurations which has been quantified by the size of the gorge. In addition, the essential residue-residue contact dynamics that contribute to the collective motion of the protein have been identified using the CAMERRA method. A critical region for gating ($\alpha 5$) tends to form an α helix under high pressure. This seems to be essential for the open-gate conformation. However, this

pressure mediated gating effect is a non-monotonous process, as we found that at extremely high pressures (p >1000 MPa), the dynamics of conformational exploration of the enzyme are restricted while the protein transforms to a closed state with an ordered α5. This study has yielded some encouraging results that correlate well with experimental evidence pressure activation. Also, it demonstrates that the CAMERRA method can be useful in evaluating pressure effects on protein conformations.

Acknowledgments

Acknowledgment is made to the Donors of the American Chemical Society Petroleum Research Fund for partial support of this research. Q.R.J. was supported by an NSF-funded graduate fellowship program pen SCALE-IT. We also acknowledge the computational support provided by the UT-ORNL Center for Molecular Biophysics and by allocations of advanced computing resources (Kraken at NICS and STAMPEDE at TACC) provided by the NSF.

5.3 Appendix 3

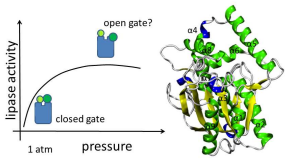

Figure 5.1: Pressure mediated activity of the PAL

(a) Unlike most enzymes, lipase activity can increase with increasing pressure as depicted here. The caricature of the corresponding structural changes due to high pressure is also shown. (b) A canonical view of the lipase with some structural elements labeled for a point of reference.

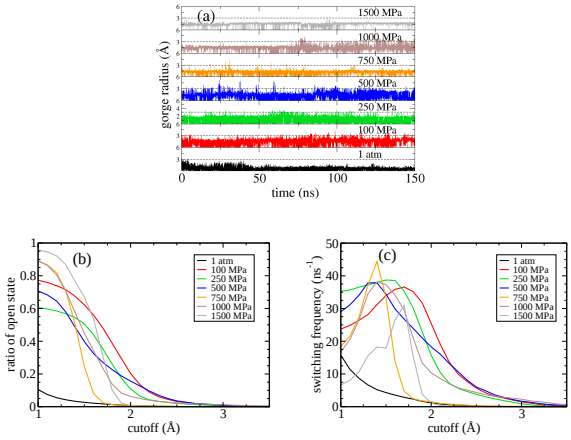

Figure 5.2: Dynamic and average gorge radius properties

The gorge radius of the lipase is shown as a function of time in (a). The color scheme representing the seven pressure systems is consistent throughout this work. The ratio of open configurations throughout the simulation is shown as a function of the gorge radius cutoff in (b). The corresponding frequency of which the lipase switches between the open and closed states is shown in (c).

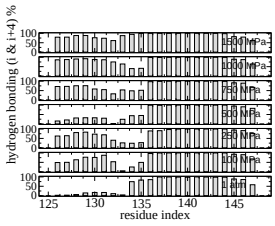

Figure 5.3: Average hydrogen bonding for lipase under various pressures

Percentages of backbone H-bonds formed between residues i (C=O) and i + 4 (N-H) of α5.

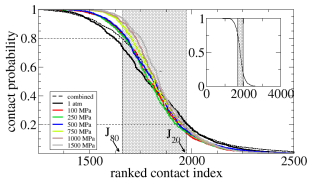

Figure 5.4: Ranked contact curve for lipase under various pressures

The ranked contact curves (RCCs) display the probability of residue-residue contacts for each individual system (solid lines) and the combined data (dashes). The contacts between J_{20} and J_{80} are selected for further contact PCA analysis. The inset presents the overall view of the combined RCC.

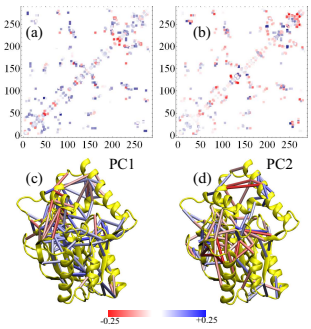

Figure 5.5: PCA contact map of lipase residue-residue interactions

Eigenvectors PC1 (a) and PC2 (b) are shown on the residue-residue contact map. The corresponding three-dimensional representation of eigenvectors PC1 and PC2 are shown as colored cylinders in (c) and (d) respectively.

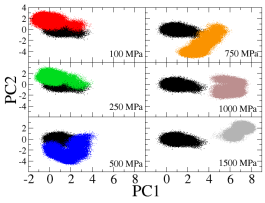

Figure 5.6: Two-dimensional contact PCA of lipase

The configurations of the lipase are projected onto PC1 and PC2. Each point represents a specific configuration of lipase obtained from a snapshot. The ensemble of room pressure is shown in each panel as a reference.

Chapter 6

Mapping Allostery through Computational Glycine Scanning and Correlation Analysis of Residue-Residue Contacts

Mapping Allostery through Computational Glycine Scanning and Correlation Analysis of Residue-residue Contacts

Quentin R. Johnson[1,2], Richard J. Lindsay[2,3], Ricky B. Nellas[4], Elias J. Fernandez[3], and Tongye Shen[1,3]*

[1]UT-ORNL Graduate School of Genome Science and Technology, Knoxville, TN, 37996

[2]Center for Molecular Biophysics, Oak Ridge National Lab, Oak Ridge, TN, 37830

[3]Department of Biochemistry and Cellular & Molecular Biology,

University of Tennessee, Knoxville, TN, 37996

[4]Institute of Chemistry, University of the Philippines Diliman, Quezon City, 1101, Philippines

Abstract

Understanding allosteric mechanisms is essential for the physical control of molecular switches targeting downstream cellular responses. However, it is difficult to decode essential allosteric motions in a high-throughput scheme. A general two-pronged approach is presented here to perform automatic data reduction of simulation trajectories. The first step involves coarse-graining and identifying the most dynamic residue-residue contacts. The second step is to perform principal component analysis on these contacts and to extract the large scale collective motions expressed via these residue-residue contacts. We demonstrated the method using

a protein complex of nuclear receptors. Using atomistic modeling and simulation, we examined the protein complex and a set of 18 glycine point mutants of residues which constitute the binding pocket of the ligand effector. The important motions that are responsible for the allostery are reported. In contrast to conventional induced-fit and lock-and-key binding mechanisms, a novel "frustrated-fit" binding mechanism of RXR for allosteric control was revealed.

Many biomolecules, especially signaling proteins, may have more than one accessible low free-energy conformational state under physiological conditions. [1, 127] The barrier(s) separating these states is not prohibitively high. [128,129] Thus, these macromolecules are able to switch their conformations upon sensing environmental perturbations. There are a variety of such perturbations, such as interaction with another macromolecule, [1, 130] interaction with a ligand, [131] solvent conditions [113, 132–134](polarity and pH), physical environment [135–139] (salt, pressure, temperature, and light), or post-translational modifications. [126, 137]

The conformational changes induced by these perturbations serve important biological functions. The changes can be local (centered around a few residues) or global (extensive and convoluted, involving the rearrangement of the whole molecule). The amplitude of these changes can range from subtle RMSD changes to partial unfolding. One of the most well-studied cases of conformational switches is allostery. [1, 140, 141] A classic example of allostery involves the cooperative or anti-cooperative affinity of ligand-binding pockets, i.e., binding of one ligand (effector) at one pocket may induce conformational changes which promote or inhibit the binding of the second ligand at a second pocket. [142]

Many successes have been achieved in understanding allostery over the last century, [143] but questions remain on the mechanistic action of allostery, especially on negative allostery, i.e., how one ligand can negatively influence a remote binding

or active site. [144, 145] Although there are textbook examples of thoroughly studied positive allosteric systems such as hemoglobin, negative allosteric regulation is infrequently reported. One possibility is that negative allostery is rarely used in nature's design of biochemical pathways. Another reason could be that identifying negative allostery with conventional methods is difficult. Indeed, allosteric mechanisms can have a variety of themes, and a specific mechanism can be painstaking to decode. [146] Given a particular state of a protein (or protein complex) that possesses multiple binding sites, can one predict whether the binding is cooperative or anticooperative? Do all negative (or positive) allosteric mechanisms share an underlying theme? Even more interestingly, some allosteric proteins (such as the one we will examine below) can display both positive or negative allostery, depending on the allosteric target. How do such molecular switches function? The answers to these questions could lead to future bioengineering and design of a new class of signaling proteins.

Several methods have been used to examine the intricate regulation and remote communication between binding sites. [131,147–151] Most methods so far have focused on the changes of elements of allostery (such as residue nativeness or residue-residue contacts) at the "mean-field" level. Typically, a group of potentially important elements of allostery are first identified either through experimental mutation or through computational study. Then, further analysis using network or phylogenetic

tree-type classification is applied to reconnect these isolated elements. [152] Since the process of identifying important elements is often performed and recorded independently, such as alanine scanning of individual residues, the connection between these isolated elements is ignored. What is missing from these approaches is the dynamic correlation between these elements. Indeed, as pointed out by bioinformatics approaches, this aspect can be examined by analyzing the correlation between these elements, i.e., the coevolution of protein residues. [150, 153] Similarly, coarse-grained models of proteins have been studied computationally to examine the local folding states (whether a residue is native or unfolded) and how the "melting" state of one residue affects another. [147, 148] In such models, the element of allostery is the physical status or chemical identity of each residue.

In this study, we address the mechanism of allostery from the viewpoint of contact events (between residues) collected from atomistic simulations. Thus, the elements of allostery are not the status of individual residues, but the status of contacts (residue-residue interactions). The focus will be the dynamic correlation between the contact events, i.e., when residues i and j form contact, whether residues k and l also form contact. We first construct atomistic models and obtain simulation trajectories. We then extract the discrete contact degrees of freedom (DOFs) from simulation snapshots via coarse-graining and identify the important dynamic contacts. Finally, a statistical method (principal component analysis) is applied to these

contact DOFs to identify the allosteric mechanism.

The method developed here is applied to a heterodimeric nuclear receptor (NR) complex. The nuclear receptor superfamily is a well-known group of signaling proteins that is responsible for sensing the presence of important small molecules for the downstream events of transcription. As the common subunit of heterodimers of the NR superfamily, promiscuous protein retinoid X receptor (RXR) displays essential allosteric regulation of other members in the superfamily. Previous experimental studies showed that the ligand binding status of RXR influences the ligand binding status of the thyroid hormone receptor (TR), [154] the constitutive androstane receptor (CAR), [155] and the liver X receptor (LXR). [156] For instance, the TR:RXR heterodimer, can only achieve full activity upon binding of the 3,3',5 triiodo-L-thyronine (T3) ligand to TR, and transactivation levels drop in the presence of the RXR ligand 9-cis retinoic acid (9C); a clear case of negative cooperativity between the two binding sites. [154] On the other hand, RXR positively regulates the binding activity of CAR. Thus, how this promiscuous protein RXR can have different functions is quite intriguing. An illustration of the TR:RXR complex (ligand-binding domain) is shown in Figure 6.1. Below, we will report how binding of 9C will induce the release of T3. We also will report an interesting "frustrated-fit" between the binding pocket and its endogenous ligand, 9C.

Systems and Methods

Atomistic Modeling of TR:RXR Complex and Glycine Scanning

Besides directly investigating the system of the wild-type protein complex TR(T3):RXR(9C), a series of single residue glycine mutants were constructed to enhance the sampling of conformational space. Traditionally, alanine scanning is used as a technique to determine the contributions of individual amino acids to a protein's structure, function, and binding activity. [157] Alanine is an ideal residue to minimize native residue-residue interactions for experimental methods, as it is relatively inert and has minimal steric and electrostatic impact. [158] In this study, a more drastic glycine mutation is used to scan the binding pocket. Glycine residues are completely devoid of sidechain and thus eliminate most residue-ligand contact interactions. As this is a perturbation study with the native structure as the initial state, the results are not impacted by potential ill effects (such as protein instability encountered in experiments) since they are not shown in a relatively short-time all-atom simulation.

As listed in Figure 6.1 and Supporting Information (SI) Table S1, the 18 residues selected to be scanned form the binding site of RXR and can potentially form direct interactions with ligand 9-cis-retinoic acid (9C). The selection was made by directly inspecting the crystal structure and representative snapshots of the simulation of the

wild-type protein. This set of 18 glycine mutants can perturb interactions between RXR and its ligand, 9C. These sidechain-deleting mutants provide an ensemble of perturbed ligand-docked (holo) structures that are closer to the configurations of apo systems. Note, however, this particular residue-ligand interaction may not be completely removed by these point mutants. For example, in mutant A327G, glycine still can hold a backbone hydrogen bond with the carboxyl end of the ligand 9C.

We also reason that, rather than comparing the snapshots from one long-time simulation of the wild type alone, an ensemble of glycine mutants will sample the representative populated configurations more efficiently in a given short time. Here we demonstrate that this perturbation scheme, provides an important ensemble of configurations and utilizes them to detect the allosteric mechanism. Furthermore, statistical analysis of this perturbative ensemble of structures reveals important allosteric information for ligand binding and ligand crosstalk. In our design of the simulation procedure, we did not use the traditional method of directly studying the apo form (void of 9C) vs the ligand-bound form of the complex. The concern is that there is no crystal structure of the apo form. The structural changes between the apo form and the holo form are known to be drastic. If we start the simulation from the holo form of the structure with ligand completely removed, it would take an unrealistic time for the system to reach the relaxed apo form. Rather, we focus on

a set of more subtle perturbations of the holo form and perform short-time simulations. Thus, this scheme has the potential of being developed into a high-throughput method.

Simulation Setup

The starting structure for each of the 19 protein complexes was constructed from PDB ID: 3UVV [145] via the xleap module of AMBER 10. [19] The force fields for protein and water are AMBER ff99SB, [159] and TIP3P, respectively. Besides solvent, there are four separate entities in each system: protein TR (internal index 1-259, corresponding to standard residue index 147-405), protein RXR (internal index 260-491, corresponding to standard index 227-458), and ligand T3 (3,3',5 triiodo-L-thyronine, internal index 492) and 9C (9-cis-retinoic acid, internal index 493). We use the standard residue index below unless otherwise specified.

Nuclear receptors have a common fold, the so called α-helical sandwich, [160, 161] which is mostly constructed from 11 to 12 helices. The loop (residues 195-202) that connects the first two helices of TR and the corresponding loop for RXR (residues 247-260) were missing in the reported PDB; these regions were modeled via the ModLoop program. [162] The ligands T3 and 9C were modeled using the antechamber module of AMBER 10. [19] The charges of the ligands were obtained from previous results [163] or from ab-initio calculations. [164]

The size of each system is roughly that of the wild type (61,625 atoms, including 17 sodium ions and 17,906 waters). Note that each system size can vary slightly due to glycine mutation, but all the systems have a box size of about $88\times98\times94$ $\AA^3$. The energy-minimized wild-type complex was used as the starting structure for every mutant studied here. From this structure, 18 single point glycine mutant complexes were constructed with the xleap module.

For each system, minimization, heating, and equilibration (for approximately 1 ns) was conducted and followed by a 20 ns NPT production run at 1 bar and 300 K using the NAMD2.7 [59] package. The Langevin thermostat was used to regulate temperature, long-range interactions were treated using the particle mesh Ewald method, [123] and all bonds involving hydrogen atoms were constrained using the SHAKE [24] algorithm. A 2 fs time step was used for the simulation with snapshots taken and trajectories generated every 1 ps, producing 20,000 snapshots for each system (and a total of 380,000 snapshots). These snapshots, containing Cartesian coordinates of all the atoms, are the starting data sets for expressing the conformations in terms of contact degrees of freedom.

From the simulation snapshots, we have applied a computer program CAMERRA (Computation of Allosteric Mechanism by Evaluating Residue-Residue Associations) to perform the analysis of the allosteric motions which we report below. The program CAMERRA with sample data and results is available upon request.

Results and Discussion

6.1 Selecting Important Contact Degrees of Freedom for Principal Component Analysis

To simplify the multidimensionality of the configurational space and to extract the allosteric motion, a data reduction scheme is desired to locate important degrees of freedom (DOFs). Often, principal component analysis is used for such tasks using biological systems. [165] Traditionally, Cartesian coordinates are used as the inputs for principal component analysis (PCA) in biomolecular simulations. [166] PCs of cartesian coordinates resemble soft "vibrational" modes of atom positions. However, protein motion is extremely complex [127] and expressing structural data in other formats, such as internal DOFs [167] and contact DOFs [125, 126] can have some advantage, depending on the type of motions (such as backbone rotation, vibration, and folding). In the current case, allostery will be examined from the viewpoint of residue-residue interaction. Thus, the PC modes reported here correspond to the "vibrational" motion of contacts (contact forming and breaking). Residue-residue contact DOFs are convenient for describing large-scale motions of proteins such as protein folding, [1] and these may be an efficient way of coarse-graining atomistic information of large biomolecular systems as well.

First, certain DOFs are removed from the system to produce coarse-grained representations of the residue level contact interactions. Here, each snapshot, initially expressed as the Cartesian coordinates of all the atoms, is condensed down to the information of whether there is any direct contact between residues or not. We use discrete values $u_{ij}= 1$ and $= 0$, respectively, to indicate whether contacts are formed. A contact is deemed to be formed between two residues when the distance between any atoms of said residues is less than 4.2 Å; in this case $u_{ij}= 1$, otherwise, $u_{ij}= 0$.

A protein (complex) of N residues has residue-residue contacts at the order of $O(N^2)$ which is typically larger than the number of cartesian DOFs, even expressed at atomic resolution. Thus, one needs more numbers to express a particular configuration using contact DOFs. Fortunately, a protein has a much more sparse contact matrix if confined to a particular well-folded basin. If each residue has a maximum m possible native contacts (practically, m is a small number at the order of 10), there are $m \times N$ total contact DOFs. Besides those contacts that are never formed, the contact DOFs that are always formed should also be excluded. Neither of these two types of contact DOFs will be able to contribute to the important motions of the protein complex or reveal the intricate interactions between contact formations. For comparison, the luxury of limiting the contact DOFs to that of the native basin was absent in a previous study on folding and structure prediction. [125]

We first examined all potential contacts and ranked the percentage of formations during the simulation. Figure 6.2 shows the ranked contact curves (RCC), defined as the individual contact probability $p(i)$ sorted by its rank i. Note that all contacts (including the N diagonal portion of the contact matrix which gives a constant 1) are included in the RCC. The x-axis value J_{50} at which the curve intercepts with y=0.5 indicates the number of contacts that are formed 50% or more during the simulation, i.e., p(J_{50})= 0.5.

Based on the sigmoidal nature of the RCCs, a sharp "transition zone" separates a left region where contacts are always formed (near 100%) and a right one where contacts were never formed during the simulation. The most interesting contact DOFs are the contacts in the transition zone. Since these contacts are not always formed, their motions will be most important in describing the conformational dynamics of the protein complex. The combined RCC, which is constructed from all snapshots, is also displayed here. We then select the interesting contact DOFs, defined as those formed between 20% and 80% of the combined RCC for the analysis below. For this system, the "20/80" cutoff leaves a manageable amount of DOFs to perform the second stage of data reduction analysis, i.e., PCA studies. The total contact DOFs used in PCA is $J \equiv J_{20} - J_{80}$. As illustrated in Figure 6.2, the combined RCC provides the cutoff values $J_{20} = 3386$ and $J_{80} = 2897$. We generate a total 489 of DOFs shown as colored cylinders in Figure 6.2b. Note that they are

broadly distributed across the complex.

It is important to point out the RCC of the combined data, i.e., the collection of all the snapshots from 18 glycine mutants, is not a direct average of individual RCC curves. We observe that the combined RCC has a more gradual decay with increasing contact index. The combined RCC is softened with a broader tail and less dramatic drop. The longer tail of the combined RCC is due to the variety of contacts that are formed in the combined ensemble. A highly formed (near 100%) contact in a particular ensemble is not guaranteed to form in all other ensembles. Thus, the combined RCC is below the individual RCCs at this region.

Besides being used to select the important contacts to include in the analysis, RCCs can also provide information on the level of compactness of each mutant. Roughly speaking, an RCC shift towards the right and up indicates more contacts are formed and thus it indicates the protein is more compact. Interestingly, all mutants show certain enhancement of contact formation compared to the wild-type complex, as shown in Figure 6.2c. This may indicate that the wild-type complex has certain frustration due to steric interactions which are alleviated by glycine mutation. One can use J_{50} as a measurement of the compactness of a structure. Another way to measure the level of contact formation is evaluating the mean contacts formed per residue. Here, each residue has about 10 contacting neighbors (including sequential neighbors) on average.

The J_{50} and mean contact per residue are strongly correlated. Both of these properties indicate an intriguing fact that the binding pocket of each mutant is able to accommodate the endogenous ligand better and makes the complex more compact in comparison to the wild type. For the extreme case, the removal of residue H435 can increase the global contacts by about 1% relative to the wild type. This indicates that the binding of ligand 9C creates steric frustration (especially at the ring end of the ligand) which forces the molecule to adopt an ensemble of more flexible conformations. Thus, the binding pocket of T3 does not form well and decreases the affinity for T3, which is consistent with the experimental evidence that upon binding 9C, RXR changes conformation and TR becomes more "loose" and dynamic. [154] Experimental evidence showed a more dispersed electron density in the holo system (with both ligands) than in the apo system (without 9C). [154] The trend is quite pronounced for TR, which leads to the conclusion that TR becomes more flexible upon the binding of 9C to RXR. [154] This conclusion is also supported by the computed radius of gyration analysis, which indicates that mutants are more compact than the wild type (data shown in SI Figure S3).

6.2 Principal Components and Projections of Conformations onto them

In the current case, PCA is performed on the selected total $J = 489$ contact DOFs introduced in the previous section. We first calculate the elements of the covariance matrix (size of J $\times$ J) $C_{\alpha\beta} \equiv C_{i:j,k:l}$ which describes the level of the correlation between the contact events u indexed by $\alpha = 1, 2, 3, \ldots, J$ (between residues i and j) with the event $\beta = 1, 2, 3, \ldots, J$ (between residues k and l). Thus, the covariance matrix C is defined as$C_{ijkl} = \langle (u_{ij} - \langle u_{ij} \rangle) \times (u_{kl} - \langle u_{kl} \rangle) \rangle$. We use a collection of snapshots (380,000 frames) from 19 mutant and wild-type simulations. The results presented below are based on this combined covariance matrix.

For PCA, the covariance matrix C is then diagonalized, and the transformation matrix U is obtained, which yields $U^T C U = \Lambda$, where matrix U gives a recipe to linearly transform the contact DOFs to another set of DOFs, and the matrix Λ is diagonal. The significance of the transformed DOFs, termed principal components (PCs), is that they are directly ranked by their importance according to eigenvalue Λ. The top PCs (containing the largest Λs) provide the major modes of protein dynamics in terms of contacts. Ideally the fluctuation of residue-residue contact and detachment events contains only the largest few PCs for the dimension reduction, and the sorted eigenvalue Λ_i drops with index i. The 65th largest PC is less than

one-tenth of the largest. The relatively slow decay of eigenvalues suggests that quite a few modes contribute to protein motion. However, it does not diminish the importance of the top PCs. These top PCs still hold the most likely motions. Although the decrease is not extremely rapid, the top 13 PCs together contribute more than half of the whole amplitude. We will focus on PC1 and PC2 below.

We display the top two PCs (normalized eigenvectors) in Figure 6.3. Principal component 1 (PC1) is of immediate interest since it represents the largest mode of movements and will likely exhibit the most interesting information. Here, in Figure 6.3a, the normalized eigenvector $\mathrm{PC1}_{ij}$ is shown on a residue-residue contact map, and the values of these 489 contacts are color-coded. The black dashed lines demarcate the intrachain contact region and the interchain region. Figure 6.3b contains the same information on PC1, but focuses on the contacts between the protein and the ligand. Although there are a variety of positive and negative values for each contact along PC1, the overall contact with ligand T3 is positively correlated along PC1 while 9C is negatively correlated. This means that the modes of ligand-protein interaction are anticorrelated, and is an indication that when one ligand is drawn closer to its binding site, the second ligand will escape from its binding site. Our results of the negative allostery indicates the binding of T3 is weakened once 9C is bound. This is consistent with the earlier experimental study that examined the binding affinity and reported an increase in dissociation constant and a decrease in

binding constant of T3 pocket in the TR(T3):RXR(9C) complex in comparison to TR(T3):RXR(ø). [154]

Besides the 2D plot (Figure 6.3a) and 1D plot (Figure 6.3b), a cylinder representation is also used to display the information of PC1 in a 3D view (front and back), as shown in Figure 6.3d. Additional side views are presented in SI Figure S5 and S6. Here the contacts are again visualized as cylinders, similar to that of the mean contact plot in Figure 6.2b. All three representations indicate the dynamics of the PC1 mode. Overall, the most dramatic dynamics of PC1 center around the binding pocket of T3, which is remarkable given that all the mutations directly probe the 9C binding of the RXR side. Previously it was reported that electron density for the T3 binding site in TR is markedly discontinuous in the TR(T3):RXR(9C) complex. [154] In this PC1 mode, there is one main region of interchain contact. Helix H11 of RXR and H10' of TR seem to hold many of the important interchain (interfacial) contact pairs. Also, the contact between helix H6' of TR and the turn region (residues 377-383) of RXR contributes to the interfacial interaction between TR and RXR.

Similarly, Figure 6.3c is a contact map view of the PC2 value for each residue-residue contact. While PC2 may be less important than PC1, there is some useful information in this mode, which will be discussed in the PC projection below. The values in PC2 mode have smaller amplitudes, as expected.

Figure 6.4 depicts the conformation (contact states) of all systems studied projected onto the top two principal components (presented in Figure 6.3). Projecting the dynamic snapshots of all systems onto these two PCs may enhance our understanding of the propagation of the allosteric control utilized by TR:RXR. In the current case, PC1 distinguishes the wild type from the mutants, as the wild type is located on the extreme positive end of PC1 in Figure 6.4, reaching values of well above 4.0. Opposite to the wild type on the extreme negative end of PC1 lies mutant R316, reaching values of less than -4.0.

It is interesting to point out mutation R316 is expected to cause the most drastic changes to the binding interaction, since the arginine to glycine mutation removes the salt-bridge interaction between the sidechain of arginine and the carboxylic group of ligand 9C. We suggest that PC1 resembles the docking mode of the carboxyl end of ligand 9C. Additionally, L433 and L436 have similar values along PC1 and PC2. These two residues have the same chemical properties, and are also found only three residues apart, which explains their similar behavior along these modes. Interestingly, A327 and R316, which are found in a very similar locale have very different PC1 values. This can be explained by the difference in their chemical properties. Compared to the mutations near the ring of 9C, those residues near the carboxyl end of 9C tend to have a lower PC1 value.

Wild-type TR:RXR shows relatively high values of PC2 as well. PC2 is used

here as another way to differentiate mutant versus wild-type conformations. Mutant Q275 shows the lowest values along PC2, while F313 shows the highest values. Mutants C432, L433, H435, and L436 (all similar locale) show very similar PC2 values. One may wonder whether 20 ns simulations are long enough to obtain convergent PCA results. To examine the dynamics aspect of protein motion projected onto the top PCs, we split the same data to two halves (10 ns each) as shown in SI Fig. S8. We can see that most systems are relaxed to a stationary ensemble and show little change between the two halves. There are a couple exceptions that indicate transient motions. Note that our perturbation scheme aims to generate a set of diverse and slightly perturbed structures, not necessarily fully equilibrated ensembles.

Conclusions

We present a method that isolates important allosteric modes from glycine mutant simulations and by performing PCA on selected dynamic contacts. This two-stage data reduction scheme ((1) selecting dynamic contacts based on RCC, and (2) rendering important contact collective modes by PCA) is efficient in terms of computational cost and still maintains the atomistic interactions of the system. The method is applied to reveal the interactions between ligands of the NR complex

TR:RXR. Ligand 9C binds RXR, which invokes frustration of the binding site, and this frustration is propagated through the interface of the TR:RXR complex. This eventually leads to a decrease in contact between TR and its ligand T3. Understanding the allostery of RXR with its endogenous ligand will serve as a model system for studying rational drug design of antagonists that control the regulation pathway. Besides other NRs in the superfamily, the method presented here can be generally useful in the study of many other allosteric systems and large-scale motions of flexible proteins.

Supporting information includes one table and eight figures. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgments

We acknowledge the computational support provided by the UT-ORNL Center for Molecular Biophysics and by allocations of advanced computing resources (Kraken at NICS and STAMPEDE at TACC) provided by the NSF.

6.3 Appendix 4

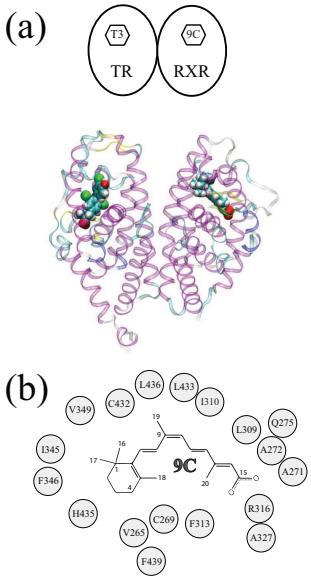

Figure 6.1: Several depictions of the RXR:TR structure of varying detail

(a) A cartoon rendering of the structure of the RXR(9C):TR(T3) complex. (b) The interaction between the RXR binding pocket and ligand 9C. The amino acid residues used to construct glycine mutants in this study are listed.

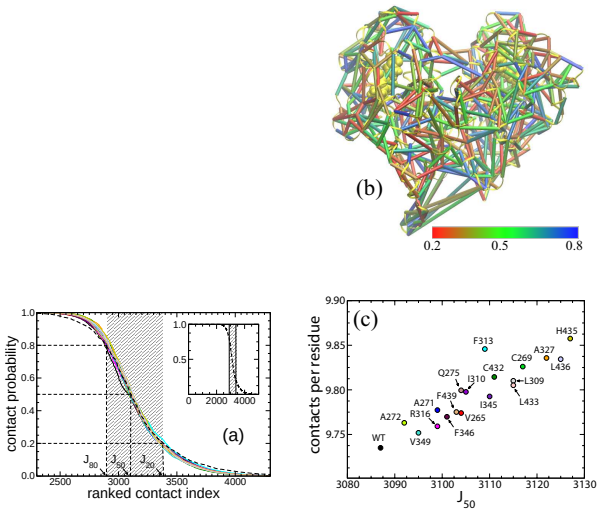

Figure 6.2: Residue-residue contact analysis of RXR-TR

(a) The ranked contact curves (RCCs) display the probability of residue-residue contacts for each individual system (thin lines) and the combined data (thick dashes). Note that there are 19 colored lines; each color represents an individual system from this study. The color scheme is displayed in (c). Positions of J_{20}, J_{50}, J_{80} of the combined RCC are labeled. The inset presents the overall view of the combined RCC. (b) The average contact probabilities for those contacts selected (between 20% and 80% formed) are displayed as cylinders (total 489) between residues in the complex structure. The color spectrum is red (0.2) to green (0.5) to blue (0.8). (c) The value of J_{50} versus the corresponding contacts per residue is displayed here for individual mutants and wild type. The color scheme for the 19 systems is consistent throughout this report.

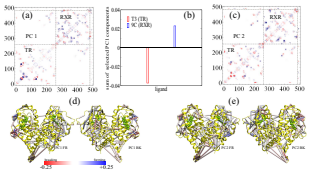

Figure 6.3: PCA contact map analysis of RXR-TR

(a) Eigenvector PC1 displayed on the residue-residue contact map. The internal index of the residues is used here for the axes. The dashed lines separates intra- and interprotein contacts. Another set of dashes demarcate the contacts formed between ligands (internal indices 492 and 493) and protein residues. (b) The sum of selected elements in eigenvector PC1 that contain the contact interaction between ligands (T3 and 9C) and the protein complex. The values of individual elements are listed in Figure S4 in SI. (c) Eigenvector PC2 displayed on the contact map. Three dimensional representations of PC1 and PC2 values are shown in (d) and (e), respectively. For clarity, only those contacts with absolute values larger than 0.03 are explicitly displayed as colored cylinders.

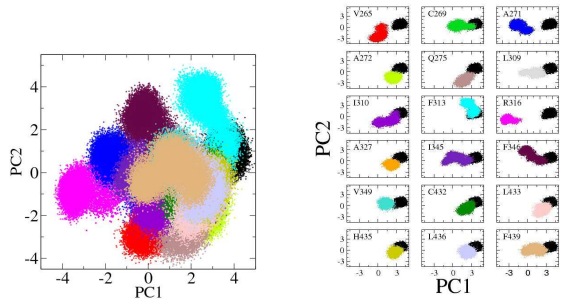

Figure 6.4: Projection of all RXR-TR snapshots onto top PCs

The configurations of the protein-ligand complex are projected onto PC1 and PC2. Each of the 380,000 points represents a specific configuration of the system obtained from a snapshot, as shown on the left. The color scheme matches previous figures. Each individual mutant is also displayed in a multi-panel view on the right for a clear comparison with the wild-type.

Chapter 7

Conclusion

Then ability of a sensing protein to switch its conformation in response to an environmental perturbation is a phenomenon that is integral to the regulation of several biochemical properties. These switching mechanisms can be quite difficult to study given the strict temporal and spatial resolution needed to provide a comprehensive view. Therefore, molecular dynamics simulation has been routinely employed in these works, as it is successful in providing these resolutions. Herein, I have presented studies that utilize MD simulation to provide insight on various swithing mechanisms. This work serves to further understanding of sensing proteins in general and provide a basis for comparison for different switching mechanisms.

In the case of the solvent dependent gating of lipase, it is apparent that the lid region is crucial to the switching mechanism to convert the lipase from a closed

to an open state. The lipase is able to maintain the open state in hydrophobic solvent and opens more in the presence of an oil-water interface. Additionally, I have shown that the gating motions of the lipase are coupled to the ordering of the $\alpha 5$ helix in both the peptide and protein form. The lipase has also been reported to display pressure regulated open-to-closed gating motions. To determine whether this switching mechanism is specific to environmental perturbation or to the lipase itself, simulations using varying degrees of pressure were conducted. It was concluded that both the solvent and pressure regulated conformational switch rely on ordering of the $\alpha 5$ region. This data suggests that different environmental factors can induce identical switching mechanisms. Additionally, switching mechanisms may be specific to certain protein subclasses. Further study is required to understand how switching mechanisms in general can be grouped together, but this is a significant finding that suggests that some general categorization is possible.

In addition, the study of the allosteric regulation of RXR has illuminated a previously undiscovered switching mechanism dubbed the "frustrated fit" mechanism. Here the CAMERRA method was used to study a negative allosteric regulation of TR by RXR in a nuclear receptor complex. This residue-residue contact evaluation method has shown that a mutation of the RXR binding pocket actually causes the binding pocket to constrict. These changes in the pocket lead to a loosening of the TR binding pocket and increase in the dynamics of the TR pocket, which is

consistent with previous experimental data. Research shows that RXR is a promiscuously dimerizing protein, and can impose either a positive or negative allosteric regulation on its partner (depending on the identity of the partner nuclear receptor). In the future it would be interesting to probe whether positive regulation occurs by the same frustrated fit mechanism, this would help to categorize the nature of this switching mechanism.

Taken as a whole, sensing proteins are a vastly complex group of biomolecules and a comprehensive understanding of them will require several lifetimes of study. This work, has shown that some switching mechanisms can be categorized by the subclass of sensing protein. Additionally, several switching mechanisms that are used by different sensing proteins have been illuminated. In time, the findings presented here can be expounded upon to reveal additional switching mechanisms and further classify them.

Bibliography

[1] Fersht, A. (1999) *Structure and mechanism in protein science : A guide to enzyme catalysis and protein folding*. (W.H. Freeman, New York).

[2] Berg, J., Tymoczko, J., Stryer, L., and Gatto, G. (2012) *Biochemistry*. (W. H. Freeman and Company).

[3] McCammon, J. A. and Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids*. (Press Syndicate of the University of Cambridge, New York, NY).

[4] Karplus, M. and Weaver, D. L. (1976) Protein folding dynamics, *Nature 260*, 404–406.

[5] McCammon, J. A., Gelin, B. R., and Karplus, M. (1977) Dynamics of folded proteins, *Nature 267*, 585–590.

[6] Ptitsyn, O. B. (1987) Protein folding: hypotheses and experiments, *J. Prot. Chem. 6*, 273–293.

[7] Dobson, C. M. (1992) Unfolded proteins, compact states and molten globules, *Curr. Opin. Struct. Biol. 2*, 6–12.

[8] Sanchez, I. E. and Kiefhaber, T. (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding, *J. Mol. Biol. 325*, 367–376.

[9] Bond, C. J. (1997) Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway, *Proc. Natl Acad. Sci. USA 94*, 13409–13413.

[10] Uttatree, S., Winayanuwattikun, P., and Charoenpanich, J. (2010) Isolation and characterization of a novel thermophilic-organic solvent stable lipase from *Acinetobacter baylyi*, *Appl. Biochem. Biotechnol. 162*, 1362–76.

[11] Day, R. (2002) Increasing temperature accelerates protein unfolding without changing the pathway of unfolding, *J. Mol. Biol. 322*, 189–203.

[12] Daggett, V. and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding, *Trends Biochem. Sci. 28*, 19–26.

[13] Fersht, A. R. (1992) The folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding, *J. Mol. Biol. 224*, 771–782.

[14] Duan, Y. and Kollman, P. A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science 282*, 740–744.

[15] Fersht, A. R. (2002) On the simulation of protein folding by short time scale molecular dynamics and distributed computing, *Proc. Natl Acad. Sci. USA 99*, 14122–14125.

[16] Mackerell, A. D. (2004) Empirical force fields for biological macromolecules: Overview and issues, *J. Comp. Chem. 25*, 1584–1604.

[17] Snow, C. D., Nguyen, H., Pande, V. S., and Gruebele, M. (2002) Absolute comparison of simulated and experimental protein-folding dynamics, *Nature 420*, 102–106.

[18] Klepeis, J., Lindorff-Larsen, K., Dror, R., and Shaw, D. (2009) Long-timescale molecular dynamics simulations of protein structure and function, *Current Opinion in Structural Biology 19*, 120–127. Theory and simulation / Macromolecular assemblages.

[19] Case, D. A., Darden, T. A., Cheatham, T. E. I., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowly, M., Walker, R. C., and Zhang, W. (2008) *AMBER 10*. (University of California, San Francisco, CA).

[20] Senn, H. and Thiel, W. (2009) Qm/mm methods for biomolecular systems, *Angewandte Chemie International Edition 48*, 1198–1229.

[21] Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-outeirino, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2007) Glycam06: A generalizable biomolecular force field. carbohydrates, *J. Comput. Chem. 29*, 622–655.

[22] Hatcher, E. R., Guvench, O., and MacKerell, A. D. (2009) Charmm additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates, and inositol, *J. Chem. Theory Comput. 5*, 1315–1327.

[23] Verlet, L. (1967) Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules, *Phys. Rev. 159*, 98–103.

[24] Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes, *J. Comp. Phys. 23*, 327 – 341.

[25] Berendsen, H., Postma, J., DiNola, A., and Haak, J. (1984) Molecular dynamics with coupling to an external bath, *J. Chem. Phys. 81*, 3684–3690.

[26] Nosé, S. (1984) A molecular dynamics method for simulations in the canonical ensemble, *Mol. Phys. 52*, 255–268.

[27] Hoover, W. G. (1985) Canonical dynamics: equilibrium phase-space distributions, *Phys. Rev. A 31*, 1695–1697.

[28] Klibanov, A. (2001) Improving enzymes by using them in organic solvents, *Nature 409*, 241–246.

[29] Nicolas, J. P. (2003) Molecular dynamics simulation of surfactin molecules at the water-hexane interface, *Biophys. J. 85*, 1377–1391.

[30] Jaeger, K.-E. and Scheidininger, B. (1997) Bacterial lipases for biotechnological applications, *Mol. Catal. 3*, 3–12.

[31] Jaeger, K.-E., Dijkstra, B. W., and Reetz, M. T. (1999) Bacterial biocatalysts: Molecular biology, structure and biotechnological applications of lipases, *Annu. Rev. Microbiol. 53*, 315–351.

[32] Ogino, H. (2008) in *Protein adaptation in extremophiles*, eds. Siddiqui, K. S. and Thomas, T. (Nova Biomedical, New York, U.S.A.), pp. 193–236.

[33] Theil, F. (1995) Lipase-supported synthesis of biologically active compounds, *Chem. Rev. 95*, 2203–2227.

[34] Schrag, J. D. and Cygler, M. (1997) Lipases and alpha/beta hydrolase fold, *Methods Enzymol. 284*, 85–107.

[35] Jaeger, K.-E., Ransac, S., Dijkstra, B. W., Colson, C., van Heuvel, M., and O., M. (1994) Bacterial Lipases, *Microbiol. 15*, 29–63.

[36] Eltaweel, M. A., Rahman, R. N., Salleh, A. B., and Basri, M. (2005) An organic solvent-stable lipase from *Bacillus sp.* strain 42, *Annals of Microb. 55*, 187–192.

[37] Nardini, M. and Dijkstra, B. W. (1999) Alpha/beta hydrolase fold enzymes: The family keeps growing, *Curr. Opin. Struct. Biol. 9*, 732–7.

[38] Gilbert, E. J. (1993) Pseudomonas lipases: Biochemical properties and molecular cloning, *Enzyme Microb. Technol. 15*, 634–645.

[39] Trodler, P., Schmid, R., and Pleiss, J. (2009) Modeling of solvent-dependent conformational transitions in *Burkholderia cepacia* lipase, *BMC Struct. Biol. 9*, 1–13.

[40] Uppenberg, J., Hansen, M., Shamkant, P., and Jones, T. (1994) The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida antarctica*, *Structure 2*, 293–308.

[41] Nardini, M., Lang, D. A., Liebeton, K., Jaeger, K., and Dijkstra, B. W. (2000) Crystal structure of *Pseudomonas aeruginosa* lipase in the open conformation, *J. Biol. Chem. 275*, 31219–31225.

[42] Lang, D., Hofmann, B., Haalck, L., Hecht, H.-J., Spener, F., Schmid, R. D., and Schomburg, D. (1996) Crystal structure of a bacterial lipase from *Chro-*

mobacterium viscosum ATCC 6918 refined at 1.6 Å, *J. Mol. Biol. 259*, 704–717.

[43] Kim, K. K., Song, H. K., Shin, D. H., Hwang, K. Y., and Suh, S. W. (1997) The crystal structure of a triacylglycerol lipase from *Pseudomonas cepacia* reveals a highly open conformation in the absence of a bound inhibitor, *Structure 5*, 173–85.

[44] Tyndall, J. D. A., Sinchaikul, S., Fothergill-Gilmore, L. A., Taylor, P., and Walkinshaw, M. D. (2002) Crystal structure of a thermostable lipase from *Bacillus stearothermophilus* P1, *J. Mol. Biol. 323*, 859 – 869.

[45] Derewenda, U., Swenson, L., Wei, Y., Green, R., Kobos, P., Joerger, R., Haas, M., and Derewenda, Z. (1994) Conformational lability of lipases observed in the absence of an oil-water interface: Crystallographic studies of enzymes from the fungi *Humicola lanuginosa* and *Rhizopus delemar*, *J. Lipid Res. 35*, 524–534.

[46] Matsumura, H., Yamamoto, T., Mori, T., and Salleh, A. (2007) Novel cation-pi interaction revealed by crystal structure of thermoalkalophilic lipase, *Proteins 70*, 592–598.

[47] Cherukuvada, S., Seshasayee, A., Raghunathan, K., and Pennathur, G. (2005) Evidence of a double-lid movement in *Pseudomonas aeruginosa* li-

pase: Insights from molecular dynamics simulations, *PLoS Comput. Biol. 1*, 182–189.

[48] Invernizzi, G., Papaleo, E., Grandori, R., and Lotti, M. (2009) Relevance of metal ions for lipase stability: Structural rearrangements induced in the *Burkholderia glumae* lipase by calcium depletion, *J. Struct. Biol. 168*, 562–570.

[49] Branco, R., Graber, M., Denis, V., and Pleiss, J. (2009) Molecular mechanism of the hydration of *Candida antarctica* lipase B in the gas phase: Water adsorption isotherms and molecular dynamics simulations, *ChemBioChem 10*, 2913–2919.

[50] Barbe, S., Lafaquiere, V., and Andre, I. (2009) Insights into lid movements of *Burkholderia cepacia* lipase inferred from molecular dynamics simulations, *Proteins: Struct., Funct., Bioinf. 77*, 509–523.

[51] Hamid, T., Rahman, R., and Basri, M. (2009) The role of lid in protein-solvent interaction of the simulated solvent stable thermostable lipase from Bacillus strain 42 in water-solvent mixtures, *Biotechnol. & Biotechnol. EQ. 23*, 1524–1530.

[52] Karjiban, R., Rahman, M., and Wahab, H. (2009) Molecular dynamics study of the structure, flexibility and dynamics of thermostable L1 lipase at high temperatures, *Protein J. 28*, 14–23.

[53] Karjiban, R., Rahman, M., and Chor, A. (2010) On the importance of the small doamin in the thermostability of thermoalkalophilic lipases from L1 and T1: Insights from molecular dynamics simulation, *Protein Pept. Lett. 17*, 1–9.

[54] Peters, G., Toxvaerd, S., and Svendsen, A. (1997) Computational studies of the activation of lipases and the effect of a hydrophobic environment, *Protein Engineering 10*, 137–147.

[55] Jensen, M., Jensen, T., Kjaer, K., Bjornholm, T., Mouritsen, O. G., and Peters, G. H. (2002) Orientation and conformations of a lipase at an interface studied by molecualr dynamics, *Biophys. J. 83*, 98–111.

[56] Ogino, H., Katou, Y., Akagi, R., Mimitsuka, T., Hiroshima, S., Gemba, Y., Doukyu, N., Yasuda, M., Ishimi, K., and Ishikawa, H. (2007) Cloning and expression of gene, and activation of an organic solvent-stable lipase from *Pseudomonas aeruginosa* LST-03, *Extremophiles 11*, 809–17.

[57] Ogino, H., Nakagawa, S., Shinya, K., Muto, T., Fujimura, N., Yasuda, M., and Ishikawa, H. (2000) Purification and characterization of organic solvent-

stable lipase from organic solvent-tolerant *Pseudomonas aeruginosa* LST-03, *J. Biosci. Bioeng. 89*, 451–457.

[58] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926–935.

[59] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J. Comput. Chem. 26*, 1781–1802.

[60] Smart, O., Neduvelil, J., Wang, X., Wallace, B., and Sansom, M. (1996) HOLE: A program for the analysis of the pore dimensions of ion channel structural models, *Jounal of Molecular Graphics 14*, 354–360.

[61] Lee, P. and Helms, V. (2012) Identifying continuous pores in protein structures with propores by computational repositioning of gating residues, *Proteins: Stucture, Function, and Bioinformatics 80*, 421–432.

[62] Xin, Y., Gadda, G., and Hamelberg, D. (2009) The cluster of hydrophobic residues controls the entrance to the active site of choline oxidase, *Biochemistry 48*, 9599–9605.

[63] Johnson, Q., Doshi, U., Shen, T., and Hamelberg, D. (2010) Water's contribution to the energetic roughness from peptide dynamics, *J. Chem. Theory Comput. 6*, 2591–2597.

[64] Yennawar, N. H., Yennawar, H. P., and Farber, G. K. (1994) X-ray crystal structure of gamma-chymotrypsin in hexane, *Biochemistry 33*, 7326–7336.

[65] Breslow, R. (1991) Hydrophobic effects on simple organic reactions in water, *Acc. Chem. Res. 24*, 159–164.

[66] Epand, R. M. (1993) *The Amphiphatic Helix*. (CRC Press, Inc., Boca Raton, Florida).

[67] Richardson, J. S. (1985) Describing patterns of protein tertiary structure, *Methods Enzymol. 115*, 341–358.

[68] Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1982) The helical hydrophobic moment: A measure of the amphiphilicity of a helix, *Nature 299*, 371–374.

[69] Shiffer, M. and Edmunson, A. B. (1967) Use of helical wheels to represent the structures of proteins and to identify segments with helical potential, *Biophys. J. 7*, 121–135.

[70] Dunnill, P. (1968) The use of helical net-diagrams to represent protein structures, *Biophys. J. 8*, 865–875.

[71] Kini, R. M. and Evans, H. J. (1989) A common cytolytic region in myotoxins, hemolysins, cardiotoxins and antibacterial peptides, *Int. L. Pept. Protein Res. 34*, 277–286.

[72] Bernheimer, A. W. and Rudy, B. (1986) Interactions between membranes and cytolytic peptides, *Biochim. Biophys. Acta 864*, 123–141.

[73] Clements, J. A. (1957) Surface tension of lung extracts, *Proc. Soc. Exp. Biol. Med. 95*, 170–172.

[74] Pattle, R. E. (1958) Properties, function, and origin of the alveolar lining layer, *Proc. R. Soc. Lond. B 148*, 217–240.

[75] Sansom, M. S. (1991) The biophysics of peptide models of ion channels, *Prog. Biophys. Mol. Biol. 55*, 139–235.

[76] Landgraf, B., Cohen, F. E., Smith, K. A., Gadski, R., and Ciardelli, T. L. (1989) Structural significance of the C-terminal amphiphilic helix of interleukin-2, *J. Biol. Chem. 264*, 816–822.

[77] DeVries, A. L. (1984) Role of glycopeptides and peptides in inhibition of crystallization of water in polar fishes, *Phil. Trans. R. Soc. Lond. B 304*, 575–588.

[78] Duman, J. G. (1982) Insect antifreezes and ice-nucleating agents, *Cryobiology 19*, 613–627.

[79] Crisma, M., Formaggio, F., Moretto, A., and Toniolo, C. (2006) Peptide helices based on alpha-amino acids, *Biopolymers (Pept Sci) 84*, 3–12.

[80] Marechal, Y. (2007) *The Hydrogen Bond and the Water Molecule: The Physics and Chemistry of Water, Aqueous and Bio-Media*. (Elsevier, Amsterdam).

[81] Vieira-Pires, R. S. and Morais-Cabral, J. H. (2010) 3_{10}-helices in channels and other membrane proteins, *J. Gen. Physiol. 136*, 585–592.

[82] Jr., R. B. and Eliezer, D. (2003) A structural and functional role for 11-mer repeats in α-synuclein and other exchangeable lipid binding proteins, *J. Mol. Biol. 329*, 763–778.

[83] Segrest, J. P., Jones, M. K., Klon, A. E., Sheldahl, C. J., Hellinger, M., Loof, H. D., and Harvey, S. C. (1999) A detailed molecular belt model

for apolipoprotein A-I in discoidal high density lipoprotein, *J. Biol. Chem. 274*, 31755–31758.

[84] Sheinerman, F. B. and Brooks, C. L. (1995) 3_{10}-helices in peptides and proteins as studied by modified Zimm-Bragg theory, *J. Am. Chem. Soc. 117*, 10098–10103.

[85] Barlow, D. J. and Thornton, J. M. (1988) Helix geometry in proteins, *J. Mol. Biol. 201*, 601–619.

[86] Topol, I. A., Burt, S. K., Deretey, E., Tang, T. H., Perczel, A., Rashin, A., and Csizmadia, I. G. (2001) α- and 3_{10}-helix interconversion: A quantum-chemical study on polyalanine systems in the gas phase and in aqueous solvent, *J. Am. Chem. Soc. 123*, 6054–6060.

[87] Smythe, M. L., Huston, S. E., and Marshall, G. R. (1995) The molten helix: Effects of solvation on the α-helical to 3_{10}-helical transition, *J. Am. Chem. Soc. 117*, 5445–5452.

[88] Gerstein, M. and Chothia, C. H. (1991) Analysis of protein loop closure: Two types of hinges produce one motion in lactate dehydrogenase, *J. Mol. Biol. 220*, 133–149.

[89] McPhalen, C. A., Vincent, M. G., Picot, P., and Jansonius, J. N. (1992) Domain closure in mitochondrial aspartate aminotransferase, *J. Mol. Biol. 227*, 197–213.

[90] Karle, I. L. and Balaram, P. (1990) Structural characteristics of α-helical peptide molecules containing Aib residues, *Biochemistry 29*, 6747–6756.

[91] Otoda, K., Kitagawa, Y., Kimura, S., and Imanishi, Y. (1993) Chain length dependent transition of 3_{10}- to α-helix of Boc-(Ala-Aib)n-OMe, *Biopolymers 33*, 1337–1345.

[92] Karle, I. L., Sukumar, M., and Balaram, P. (1986) Parallel packing of α-helices in crystals of the zervamicin IIA analog Boc-Trp-Ile-Ala-Aib- Ile-Val-Aib-Leu-Aib-Pro-OMe.2H_2O, *Proc. Natl. Acad. Sci. U.S.A. 83*, 9284–9288.

[93] Karle, I. L., Flippen-Anderson, J. L., Sukumar, M., and Balaram, P. (1988) Monoclinic polymorph of Boc-Trp-Ile-Ala-Aib-Ile- Val-Aib-Leu-Aib-Pro-OMe(anhydrous). Parallel packing of 3_{10}-/α-helices and a transition of helix type, *Int. J. Pept. Protein Res. 31*, 567–576.

[94] Marshall, G. R., Hodgkin, E. E., Langs, D. A., Smith, G. D., Zabrocki, J., and Leplawy, M. T. (1990) Factors governing helical preference of peptides

containing multiple α,α-dialkyl amino acids, *Proc. Natl. Acad. Sci. USA 87*, 487–491.

[95] Aleman, C., Subirana, J. A., and Perez, J. J. (1992) A molecular mechanical study of the structure of poly(α-aminoisobutyric acid), *Biopolymers 32*, 621–631.

[96] Hodgkin, E. E., Clark, J. D., Miller, K. R., and Marshall, G. R. (1990) Conformational analysis and helical preferences of normal and α,α-dialkyl amino acids, *Biopolymers 30*, 533–546.

[97] Huston, S. E. and Marshall, G. R. (1994) $\alpha/3_{10}$-helix transitions in α-methylalanine homopeptides: Conformational transition pathway and potential of mean force, *Biopolymers 34*, 75–90.

[98] Bellanda, M., Mammi, S., Geremia, S., Demitri, N., Randaccio, L., Broxterman, Q. B., Kaptein, B., Pengo, P., Pasquato, L., and Scrimin, P. (2007) Solvent polarity controls the helical conformation of short peptides rich in C^{α}-tetrasubstituted amino acids, *Chem. Eur. J. 13*, 407–416.

[99] Karle, I. L., Flippen-Anderson, J. L., Gurunath, R., and Balaram, P. (1994) Facile transition between 3_{10}- and α-helix: Structures of 8-, 9-, and 10-residue peptides containing the -(Leu-Aib-Ala)2-Phe-Aib-fragment, *Prot. Sci. 3*, 1547–1555.

[100] Vijayakumar, E. K. S. and P., B. (1983) Stereochemistry of α-aminoisobutyric acid peptides in solution: Helical conformations of protected decapeptides with repeating Aib-L-Ala and Aib-L-Val sequences, *Biopolymers 22*, 2133–2140.

[101] Borhani, D. W., Rogers, D. P., Engler, J. A., and Brouillette, C. G. (1997) Crystal structure of truncated human apolipoprotein A-I suggests a lipid-bound conformation, *Proc. Natl. Acad. Sci. USA 94*, 12291–12296.

[102] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. J., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc. 117*, 5179–5197.

[103] Johnson, Q. R., Nellas, R. B., and Shen, T. (2012) Solvent-dependent gating motions of an extremophilic lipase from *Pseudomonas aeruginosa*, *Biochemistry 51*, 6238–6245.

[104] Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577–2637.

[105] Armen, R., Alonso, D. O. V., and Daggett, V. (2003) The role of α-, 3_{10}-, and π-helix in helix to coil transitions, *Prot. Sci. 12*, 1145–1157.

[106] Mikhonin, A., Bykov, S., Myshakina, N., and Asher, S. (2006) Peptide secondary structure folding reaction coordinate: Correlation between uv raman amide iii frequency, psi ramachandran angle, and hydrogen bonding, *J. Phys. Chem. B 110*, 1928–1943.

[107] Nellas, R. B., Glover, M. M., Hamelberg, D., and Shen, T. (2012) High-pressure effect on the dynamics of solvated peptides, *J. Chem. Phys. 136*, 145103.

[108] Henchman, R. H. and McCammon, J. A. (2002) Extracting hydration sites around proteins from explicit water simulations, *J. Comput. Chem. 23*, 861–869.

[109] Haynes, W. H. (2013) *CRC Handbook of Chemistry and Physics*. (CRC Press, Boca Raton, Florida), 94th edition.

[110] Reichardt, C. (2003) *Solvents and Solvent Effects in Organic Chemistry*. (Wiley, Weinheim, Germany), 3rd edition.

[111] Farese, R. V. and Walther, T. C. (2009) Lipid droplets finally get a little r-e-s-p-e-c-t, *Cell 139*, 855–860.

[112] Rehm, S., Trodler, P., and Pleiss, J. (2010) Solvent-induced lid opening in lipases: A molecular dynamics study, *Protein Sci. 19*, 2122–2130.

[113] Nellas, R. B., Johnson, Q. R., and Shen, T. (2013) Solvent-induced α- to 3_{10}-helix transition of an amphiphilic peptide, *Biochemistry 52*, 7137–7144.

[114] Eisenmenger, M. J. and Reyes-De-Corcuera, J. I. (2010) Kinetics of lipase catalyzed isoamyl acetate synthesis at high hydrostatic pressure in hexane, *Biotechnol. Lett. 32*, 1287–91.

[115] Eisenmenger, M. J. and Reyes-De-Corcuera, J. I. (2010) Enhanced synthesis of isoamyl acetate using an ionic liquid–alcohol biphasic system at high hydrostatic pressure, *J. Mol. Catal. B: Enzym. 67*, 36–40.

[116] Eisenmenger, M. J. and Reyes-De-Corcuera, J. I. (2009) High hydrostatic pressure increased stability and activity of immobilized lipase in hexane, *Enzyme Microb. Tech. 45*, 118–125.

[117] Eisenmenger, M. J. and Reyes-De-Corcuera, J. I. (2009) High pressure enhancement of enzymes: A review, *Enzyme Microb. Technol. 45*, 331–347.

[118] Hawley, S. A. (1971) Reversible pressure-temperature denaturation of chymotrypsinogen, *Biochemistry 10*, 2436–2442.

[119] Paschek, D., Hempel, S., and Garcia, A. E. (2008) Computing the stability diagram of the trp-cage miniprotein, *Proc. Natl. Acad. Sci. U.S.A. 105*, 17754–17759.

[120] Trzesniak, D., Lins, R. D., and van Gunsteren, W. F. (2006), *Proteins: Struct., Funct., Bioinf. 65*, 136–144.

[121] Kato, C., Sato, T., Abe, F., Ohmae, E., Tamegai, H., Nakasone, K., Siddiqui, K. S., and Thomas, T. (2008) in *Protein Adaptation in Extremophiles*, eds. Siddiqui, K. S. and Thomas, T. (Nova Biomedical, New York, U.S.A.), pp. 167–191.

[122] Balny, C., Masson, P., and Heremans, K. (2002) High pressure effects on biological macromolecules: from structural changes to alteration of cellular processes, *Biochim. Biophys. Acta 1595*, 3–10.

[123] Darden, T., York, D., and Pedersen, L. (1993) Particle mesh ewald: An n.log(n) method for ewald sums in large systems, *J. Chem. Phys. 98*, 10089–10092.

[124] Johnson, Q. R., Lindsay, R. J., Nellas, R. B., Fernandez, E. J., and Shen, T. (2015) Mapping allostery through computational glycine scanning and correlation analysis of residue-residue contacts, *Biochemistry*.

[125] Shen, T., Zong, C., Hamelberg, D., McCammon, J. A., and Wolynes, P. G. (2005) The folding energy landscape and phosphorylation: Modeling the conformational switch of the NFAT regulatory domain, *FASEB J. 19*, 1389–1395.

[126] Latzer, J., Shen, T., and Wolynes, P. G. (2008) Conformational switching upon phosphorylation: A predictive framework based on energy landscape principles, *Biochemistry 47*, 2110–2122.

[127] Frauenfelder, H., Deisenhofer, J., and Wolynes, P. G. (2000) *Simplicity and Complexity in Proteins and Nucleic Acids*. (Dahlem University Press, Berlin).

[128] Miyashita, O., Onuchic, J. N., and Wolynes, P. G. (2003) Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins, *Proc. Natl. Acad. Sci. U.S.A. 100*, 12570–12575.

[129] Tripathi, S. and Portman, J. J. (2013) Allostery and folding of the n-terminal receiver domain of protein ntrc, *J. Phys. Chem. B 117*, 13182–13193.

[130] Taylor, M. E. and Drickamer, K. (2003) *Introduction to Glycobiology*. (Oxford University Press, New York).

[131] Hardy, J. A. and Wells, J. A. (2004) Searching for new allosteric sites in enzymes, *Curr. Opin. Struct. Biol. 14*, 706–715.

[132] Colombo, M. F., Rau, D. C., and Parsegian, V. A. (1992) Protein solvation in allosteric regulation: a water effect on hemoglobin, *Science 256*, 655–659.

[133] Ehrlich, L. S., Liu, T., Scarlata, S., Chu, B., and Carter, C. A. (2001) HIV-1 capsid protein forms spherical (immature-like) and tubular (mature-like) particles in vitro: structure switching by ph-induced conformational changes, *Biophys. J. 81*, 586–594.

[134] Barrera, F. N., Fendos, J., and Engelman, D. M. (2012) Membrane physical properties influence transmembrane helix formation, *Proc. Natl. Acad. Sci. U.S.A. 109*, 14422–14427.

[135] Wroblowski, B., Díaz, J. F., Heremans, K., and Engelborghs, Y. (1996) Molecular mechanisms of pressure induced conformational changes in bpti, *Proteins: Structure, Function, and Bioinformatics 25*, 446–455.

[136] Schlesinger, M. J. (1990) Heat shock proteins, *J. Biol. Chem. 265*, 12111–12114.

[137] Ha, J. and Loh, S. N. (1990) Protein conformational switches: From nature to design, *Chemistry 18*, 7984–7999.

[138] Imamoto, Y., Kamikubo, H., Harigai, M., Shimizu, N., and Kataoka, M. (2002) Light-induced global conformational change of photoactive yellow protein in solution., *Biochemistry 41*, 13595–13601.

[139] Vreede1, J., Juraszek, J., and Bolhuis, P. G. (2010) Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein, *Proc. Natl. Acad. Sci. U.S.A. 107*, 2397–2402.

[140] Hill, T. L. (2011) *Cooperativity Theory in Biochemistry: Steady-State and Equilibrium Systems*. (Springer), First edition.

[141] Cui, Q. and Karplus, M. (2008) Allostery and cooperativity revisited, *Protein Science 17*, 1295–1307.

[142] Pauling, L. (1935) The oxygen equilibrium of hemoglobin and its structural interpretation, *Proc. Natl. Acad. of Sci. U.S.A. 21*, 186–191.

[143] Hill, R. (1926) Chemical nature of haemochromogen and its carbon monoxide compound, *Proc. R. Soc. B 100*, 419–430.

[144] Stevens, S., Sanker, S., Kent, C., and Zuiderweg, E. (2001) Delineation of the allosteric mechanism of a cytidylyltransferase exhibiting negative cooperativity, *Nat. Struct. Biol. 8*, 947–952.

[145] Putcha, B. K., Wright, E., Brunzelle, J. S., and Fernandez, E. J. (2012) Structural basis for negative cooperativity within agonist-bound TR:RXR heterodimers, *Proc. Natl. Acad. of Sci. U.S.A. 109*, 6084–6087.

[146] Ferreon, A. C., Ferreon, J. C., Wright, P. E., and Deniz, A. A. (2013) Modulation of allostery by protein intrinsic disorder, *Nature 498*, 390–394.

[147] Pan, H., Lee, J. C., and Hilser, V. J. (2000) Binding sites in escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble, *Proc. Natl. Acad. Sci. U.S.A. 97*, 12020–12025.

[148] Freire, E. (1999) The propagation of binding interactions to remote sites in proteins: Analysis of the binding of the monoclonal antibody d1.3 to lysozyme, *Proc. Natl. Acad. Sci. U.S.A. 96*, 10118–10122.

[149] Sethy, A., Tian, J., Derdeyn, C. A., Korber, B., and Gnanakaran, S. (2013) A mechanistic understanding of allosteric immune escape pathways in the HIV-1 envelope glycoprotein., *PLoS Comput. Biol. 9*, e1003046.

[150] Shulman, A. I., Larson, C., Mangelsdorf, D. J., and Ranganathan, R. (2004) Structural determinants of allosteric ligand activation in rxr heterodimers., *Cell. 116*, 417–429.

[151] Jackson, M. B. (2006) *Molecular and Cellular Biophysics*. (Cambridge University Press), 1st edition.

[152] Newman, M. (2010) *Networks: an introduction*. (Oxford University Press), First edition.

[153] Lockless, S. W. and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families, *Science 286*, 295–299.

[154] Putcha, B.-D. K. and Fernandez, E. J. (2009) Direct inter-domain interactions can mediate allosterism in the thyroid receptor, *J. Biol. Chem. 284*, 22517–22524.

[155] Wright, E., Busby, S. A., Wisecarver, S., Vincent, J., Griffin, P. R., and Fernandez, E. J. (2011) Helix 11 dynamics is critical for constitutive androstane receptor activity, *Structure 19*, 37–44.

[156] Shulman, A. I., Larson, C., Mangelsdorf, D. J., and Ranganathan, R. (2004) Structural determinants of allosteric ligand activation in RXR heterodimers, *Cell 116*, 417–429.

[157] Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2007) Hot spots-a review of the protein-protein interface determinant amino-acid residues, *Proteins: Structure, Function, and Bioinformatics 68*, 803–812.

[158] DeLano, W. L. (2002) Unraveling hot spots in binding interfaces: progress and challenges., *Curr. Opin. Struct. Biol. 12*, 14–20.

[159] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple amber force fields and development

of improved protein backbone parameters, *Proteins: Structure, Function, and Bioinformatics 65*, 712–725.

[160] Moras, D. and Gronemeyer, H. (1998) The nuclear receptor ligand-binding domain: structure and function, *Curr. Opin. Cell Biol. 10*, 384–391.

[161] Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H., and Moras, D. (1995) Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-α, *Nature 375*, 377–382.

[162] Fiser, A. and Sali, A. (2003) Modloop: automated modeling of loops in protein structures, *Bioinformatics. 19*, 2500–2501.

[163] Martinez, L., Webb, P., Polikarpov, I., and Skaf, M. S. (2006) Molecular Dynamics Simulations of Ligand Dissociation from Thyroid Hormone Receptors: Evidence of the Likeliest Escape Pathway and Its Implications for the Design of Novel Ligands, *J. Med. Chem. 49*, 23–26.

[164] Valiev, M., Bylaska, E. J., Govind, N., Kowalski, K., Straatsma, T., van Dam, H. J. J., Wang, D., Nieplocha, J., Apra, E., Windus, T. L., and de Jong, W. A. (2010) NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations, *Comput. Phys. Commun. 181*, 1477–1489.

[165] Jolliffe, I. T. (1986) *Principal component analysis*. (Springer-Verlag, New York).

[166] Garcia, A. E. (1992) Large-amplitude nonlinear motions in proteins, *Phys. Rev. Lett. 68*, 2696–2699.

[167] Altis, A., Nguyen, P., Hegger, R., and Stock, G. (2007) Dihedral angle principal component analysis of molecular dynamics simulations, *J. of Chem. Phys. 126*, 244111:1–10.

Vita

Quentin Ramon Johnson was born on March 11, 1984 in Warner Robbins, GA to George and Sylvia Johnson. He had a simple childhood, playing in the countryside of scenic Jackson, GA. Early on, his athletic aptitude was evident as he excelled in football, wrestling, baseball, and basketball. However, it was his mind rather than his body that grasped the attention of the masses. When he came to age, he attended Eagles Landing High School where he flourished academically and became interested in science, thanks to his physics teacher and mentor, Mr. Lee. Upon graduation from high school, he enrolled in Georgia State University where he would eventually earn a Bachelor of Science degree in Chemistry and a Master of Science degree in Computational Chemistry. During his pursuit of the BS degree, Quentin pledged his allegiance to Omega Psi Phi Fraternity Incorporated. This decision would prove to be one of the most important of his life, for this experience taught him the virtues of manhood, humbleness, and uplift. After graduating from GSU, he was advised by the wise, yet relatable Dr. Kennedy to pursue a Masters degree.

Holding Dr. Kennedy's advice most precious, he enrolled in the Masters program at the Department of Chemistry at GSU. At this time, Dr. Kennedy also introduced Quentin to Dr. Hamelberg, who would become his thesis adviser. With the help of Dr. Hamelberg our hero quickly completed his MS and set his aspirations even higher, he was determined to earn a PhD. So, Quentin traveled to Knoxville, TN to begin his Doctoral journey. While there, he found a proverbial diamond in the rough, Esther. He immediately became enamored with this pillar of beauty, and set out to gain her affections. Their romance was one of epic passion that led to marriage and procreation. Their first born child arrived in this world on October 21, 2013 with good spirits and raven mane. Rejuvenated by the love of his wife and the pride he felt for his progeny, Quentin trudged through the drudgery of his PhD work and kept his sights on the light at the end of the tunnel. Then, on the day of June 18, 2015, Quentin defended his dissertation entitled, Elucidation of Conformational Switching Mechanisms of Sensing Proteins by Molecular Dynamics Perturbation Studies, earning his PhD in Life Sciences from the Graduate School of Genome Science and Technology. Currently, Dr. Quentin Johnson is a postdoctoral researcher for NIMBIOS where he continues the research that is discussed herein.